

STIC-Biot ch/Ch mLib

49689

From: Davis, Natalie  
Sent: Monday, August 27, 2001 2:29 PM  
To: STIC-Biotech/ChemLib  
Subject: 09/589777

Please search SEQ ID NO:2

*Natalie A. Davis, PhD*

Patent Examiner

Art Unit 1642

CM1 Rm 8B13

ph (703) 308-6410

Point of Contact:  
Toby Port  
Technical Info. Specialist  
**CM1 1E01 TEL: 308-3534**

RECEIVED  
AUG 28 2001  
STIC

**THIS PAGE BLANK (USPTO)**

**CM1 1E01 TEL: 308-3234**  
**Technical Info. Specialist**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2001, 12:49:51 : Search time 14.49 Seconds  
(without alignments)  
967.296 Million cell updates/sec

Title: US-09-589-777A-2

Perfect score: 968  
Sequence: 1 HTTHQDEQPVLLVLAALNTPLS.....CHNSYIVACIENSFMTSESK 184

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	968	100.0	1774	2 B56101	collagen alpha 1(X
2	965	99.7	1315	2 A56101	collagen alpha 1(X
3	840	86.8	684	2 A53019	collagen alpha 1(X
4	563	58.2	1388	2 A53317	collagen alpha 1(X
5	376	38.8	650	2 T22002	hypothetical prote
6	79	8.2	427	2 A53798	58k membrane-assoc
7	78	8.1	1055	2 S53597	chlorophyll a/b-bi
8	78	8.1	1955	1 ACCH	agrin precursor
9	77.5	8.0	474	2 T10011	probable penicillin
10	77.5	8.0	6420	2 T30283	polyketide synthas
11	77	8.0	208	2 T45272	methyltransferase
12	77	8.0	1575	2 S68448	synaptotagmin, 170k
13	76.5	7.9	1054	2 T30933	chitinase (EC 3.2.
14	76	7.9	309	1 E65112	hypothetical 34.6
15	76	7.9	309	2 E85985	hypothetical prote
16	75.5	7.8	183	4 S53318	hypothetical prote
17	75.5	7.8	4924	2 T50176	probable peptidase
18	75	7.7	587	1 SYHUA6	5-aminolevulinat
19	74.5	7.7	3670	2 T36249	CDA peptidase synth
20	74.5	7.7	7576	2 T17428	FK506 polyketide s
21	74	7.6	244	2 C45729	exopolysphatase
22	74	7.6	513	1 A45333	exopolysphatase
23	74	7.6	513	2 H85893	hypothetical prote
24	74	7.6	4180	2 G83559	hypothetical prote
25	73	7.5	477	2 T05202	pectinesterase hom
26	72.5	7.5	419	2 D42725	nitrile hydratase
27	72.5	7.5	442	2 I47074	gene CDS protein -
28	72.5	7.5	463	2 B65994	Xaa-His dipeptidase
29	72.5	7.5	739	1 VHWI6B	nucleocapsid prote

30	72.5	7.5	800	2 T25140	hypothetical prote
31	72	7.4	732	2 C84487	hypothetical prote
32	72	7.4	2899	2 T21546	hypothetical prote
33	71.5	7.4	204	2 B45022	CRK-I - human
34	71.5	7.4	850	2 S56015	gastric mucin MUC5
35	71.5	7.4	5232	2 A45086	HC-toxin synthetas
36	71	7.3	512	2 A70569	probable cpsA prot
37	71	7.3	587	2 J70278	5-aminolevulinat
38	71	7.3	646	2 D70939	hypothetical prote
39	71	7.3	841	2 B71212	hypothetical prote
40	71	7.3	1289	2 T18212	parasporel crystal
41	70.5	7.3	252	2 F83350	hypothetical prote
42	70.5	7.3	257	2 S24812	minor fibrillar pro
43	70.5	7.3	316	1 F64966	probable transcript
44	70.5	7.3	319	2 B35090	MHC nonclassical c
45	70.5	7.3	416	2 G86065	probable cytochrom

## ALIGNMENTS

RESULT 1  
B56101  
collagen alpha 1(XVIII) chain precursor, long splice form - mouse  
N:Contains: collagen alpha 1(XVIII) chain precursor, medium splice form; endostatin  
C:Species: Mus musculus (house mouse)  
C>Date: 03-Oct-1995 #sequence revision 08-May-1998 #text-change 31-Mar-2000  
C:Accession: B56101; C56101; S72450; S65595; P0675; A54072; A58816  
J:Rein, M.; Pihlajaniemi, T.  
J: Biol. Chem. 270, 4705-4711, 1995  
A:Title: Identification of three N-terminal ends of type XVIII collagen chains and th  
lif homologous to rat and Drosophila fibrillized proteins.  
A:Reference number: A56101; MUID:95181468  
A:Accession: B56101  
A:Molecule type: mRNA  
A:Residues: 1-562 <REH1>  
A:Cross-references: GB:U11637; NID:9618429; PIDN:AAC52179.1; PID:9618430  
A:Experimental source: splice form clone PE17.24  
A:Accession: C56101  
A:Molecule type: mRNA  
A:Residues: 1-239,487-562 <REH2>  
A:Cross-references: GB:U11637; NID:9618429  
A:Experimental source: splice form clones PE8.1, PE19, PE15.2  
R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.  
submitted to the EMBL Data Library, August 1993  
A:Reference number: S72450  
A:Accession: S72450  
A:Molecule type: mRNA  
A:Residues: 487-1146, 'L', 1148-1193, 'F', 1195-1210, 'R', 1212-1512, 'L', 1514-1522, 'F', 1524  
A:Cross-references: EMBL:L22545; NID:9348968; PIDN:AA19787.1; PID:9511298  
R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.  
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994  
A:Title: Isolation and sequencing of cDNAs for proteins with multiple domains of gly-  
A:Reference number: A58370; MUID:94240111  
A:Accession: S65595  
A:Molecule type: mRNA  
A:Residues: 487-1512, 'L', 1514-1522, 'F', 1524-1683, 'V', 1685-1774 <OH2>  
A:Cross-references: EMBL:L22545  
R:Abel, N.; Muragaki, Y.; Yoshioaka, H.; Inoue, H.; Ninomiya, Y.  
Biochem. Biophys. Res. Commun. 196, 576-582, 1993  
A:Title: Identification of a novel collagen chain represented by extensive interrupt  
A:Reference number: P0675; MUID:94059075  
A:Accession: P0675  
A:Molecule type: mRNA  
A:Residues: 635-1774 <ABE>  
R:Rein, M.; Hintikka, E.; Pihlajaniemi, T.  
J. Biol. Chem. 269, 13929-13935, 1994  
A:Title: Primary structure of the alpha chain of mouse type XVIII collagen, partial  
collagen chain.  
A:Reference number: A54072; MUID:94245707  
A:Accession: A54072  
A:Molecule type: DNA; mRNA  
A:Residues: 1293-1403, 'R', 1405-1774 <REH3>

1

A:Cross-references: GB:U03714; NID:9487733; PIDN:AAA20657.1; PID:9487734  
 R:O'Reilly, M.S.; Boehm, T.; Shing, Y.; Fukui, N.; Vasios, G.; Lane, W.S.; Flynn, E.; B  
 Cell 88, 277-285, 1997  
 A:Title: Endostatin: an endogenous inhibitor of angiogenesis and tumor growth.  
 A:Reference number: A58816; M0ID:97160848  
 A:Accession: A58816  
 A:Molecule type: Protein  
 A:Residues: 1591-1610 <ORE>  
 A:Experimental source: hemangioendothelium cells  
 C:Comment: Inhibits endothelial cell proliferation  
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
 C:Comment: The different splice forms of collagen alpha 1(XVII) may be involved in per  
 C:Comment: Endostatin is released from collagen alpha 1(XVII) chain by the action of un  
 C:Comment: may be useful in treating solid tumors.  
 C:Genetics:  
 A:Gene: MGI:Col18a1  
 A:Cross-references: MGI:71175  
 A:Map position: 10:41.0  
 A:Intons: 1295/3; 1310/1; 1331/1; 1345/3; 1388/3; 1437/1; 1461/3; 1505/3; 1516/3; 1599/  
 A:Note: the 1st of introns is incomplete  
 C:Superfamily: unassigned collagens  
 C:Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc  
 F:1-1774/Product: collagen alpha 1(XVII) chain precursor, long splice form #status pred  
 F:1-239,487-1774/Product: collagen alpha 1(XVII) chain precursor, medium splice form #s  
 F:1-24/Domain: signal sequence #status predicted <SIG>  
 F:361-486/Region: flizzed similarity  
 F:786-812/Domain: collagenous #status predicted <CO01>  
 F:823-896/Domain: collagenous #status predicted <CO02>  
 F:921-1044/Domain: collagenous #status predicted <CO03>  
 F:1066-1148/Domain: collagenous #status predicted <CO04>  
 F:1163-1204/Domain: collagenous #status predicted <CO05>  
 F:1218-1290/Domain: collagenous #status predicted <CO06>  
 F:1301-1333/Domain: collagenous #status predicted <CO07>  
 F:1346-1389/Domain: collagenous #status predicted <CO08>  
 F:1351-1353/Region: cell attachment (R-G-D) motif  
 F:1377-1458/Domain: collagenous #status predicted <CO09>  
 F:1442-1459/Domain: collagenous #status predicted <CO10>  
 F:1591-1774/Product: endostatin #status predicted <EST>  
 F:1598-1774/Region: multiplexin collagen carboxyl-terminal similarity  
 F:354,361,947/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:699,704,1716/Binding site: carbohydrate (Ser) (covalent) #status predicted  
 F:910,913,1053/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 100.0%; Score 968; DB 2; Length 1774;  
 Best Local Similarity 100.0%; Pred. No. 3,5e-86;  
 Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HTHODFQPVHLVALNTPLSGKMGIRGADFOCFQOARAVGLSGTFRAPFLSSRLQDLXSI 60  
 |||||||  
 DB 1591 HTHODFQPVHLVALNTPLSGKMGIRGADFOCFQOARAVGLSGTFRAPFLSSRLQDLXSI 1650  
 |||||||  
 OY 61 VRRADRGSPVIVNLKDEVLSPSWDSLFSSGQGLQPGARIFSEPDGRVLRHPAMPQKSVW 120  
 |||||||  
 DB 1651 VRRADRGSPVIVNLKDEVLSPSWDSLFSSGQGLQPGARIFSEPDGRVLRHPAMPQKSVW 1710  
 |||||||  
 OY 121 HGSPSPRRRLMESYCEWRETTGATGQASSLSGRLLLEQKAASCHSYVLCLENSFMT 180  
 |||||||  
 DB 1711 HGSPSPRRRLMESYCEWRETTGATGQASSLSGRLLLEQKAASCHSYVLCLENSFMT 1770  
 |||||||  
 OY 181 SFSK 184  
 |||||  
 DB 1771 SFSK 1774

RESULT 2  
 A56101  
 N:collagen alpha 1(XVII) chain precursor, short splice form - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 03-Oct-1995 #sequence, revision 08-May-1998 #text\_change 31-Mar-2000  
 C:Accession: A56101; A58371; S72450; S65595

R:Rehm, M.; Pihlajaniemi, T.  
 J. Biol. Chem. 270, 4705-4711, 1995  
 A:Title: Identification of three N-terminal ends of type XVII collagen chains and th  
 tif homologous to rat and Drosophila frizzled proteins.  
 A:Reference number: A56101; M0ID:95181468  
 A:Accession: A56101  
 A:Molecule type: mRNA  
 A:Residues: 1-103 <REH1>  
 A:Cross-references: GB:U11636; NID:9618427; PIDN:AMC5218.1; PID:9618428  
 R:Rehm, M.; Pihlajaniemi, T.  
 Proc. Natl. Acad. Sci. U.S.A. 91, 4234-4238, 1994  
 A:Title: Alpha1(XVII), a collagen chain with frequent interruptions in the collagen  
 A:Reference number: A58371; M0ID:94240112  
 A:Accession: A58371  
 A:Molecule type: mRNA  
 A:Residues: 1-928 <REH2>  
 A:Cross-references: GB:U1698; NID:9404754; PIDN:AAA37434.1; PID:9553894  
 R:Oh, S.P.; Kamagata, Y.; Muraqaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.  
 Submitted to the EMBL Data Library, August 1993  
 A:Reference number: S72450  
 A:Accession: S72450  
 A:Molecule type: mRNA  
 A:Residues: 28-687, 'U', 689-734, 'F', 736-751, 'R', 753-1315 <OH>  
 A:Cross-references: EMBL:U22545; NID:9348968; PIDN:AAA19787.1; PID:9511298  
 R:Oh, S.P.; Kamagata, Y.; Muraqaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.  
 Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994  
 A:Title: Isolation and sequencing of cDNAs for proteins with multiple domains of gly-  
 A:Reference number: A58370; M0ID:94240111  
 A:Accession: S65595  
 A:Molecule type: mRNA  
 A:Residues: 28-1315 <OH>  
 A:Cross-references: EMBL:U22545  
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating uni  
 C:Comment: The different splice forms of collagen alpha 1(XVII) may be involved in p  
 C:Comment: Endostatin is released from collagen alpha 1(XVII) chain by the action of  
 C:Comment: may be useful in treating solid tumors.  
 C:Genetics:  
 A:Gene: MGI:Col18a1  
 A:Cross-references: MGI:71175  
 A:Map position: 10:41.0  
 C:Superfamily: unassigned collagens  
 C:Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteog  
 F:1-25/Domain: signal sequence #status predicted <SIG>  
 F:24-235/Region: thrombospondin amino-terminal similarity  
 F:26-1315/Product: collagen alpha 1(XVII) chain, short splice form #status predicted  
 F:327-353/Domain: collagenous #status predicted <CO1>  
 F:364-437/Domain: collagenous #status predicted <CO2>  
 F:462-583/Domain: collagenous #status predicted <CO3>  
 F:607-689/Domain: collagenous #status predicted <CO4>  
 F:704-745/Domain: collagenous #status predicted <CO5>  
 F:759-831/Domain: collagenous #status predicted <CO6>  
 F:842-874/Domain: collagenous #status predicted <CO7>  
 F:887-910/Domain: collagenous #status predicted <CO8>  
 F:992-994/Region: cell attachment (R-G-D) motif  
 F:918-969/Domain: collagenous #status predicted <CO9>  
 F:983-1000/Domain: collagenous #status predicted <CO10>  
 F:1132-1315/Product: endostatin #status predicted <EST>  
 F:1139-1315/Region: multiplexin collagen carboxyl-terminal similarity  
 F:126,488/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:172-228/Disulfide bonds: #status predicted  
 F:240,245,1257/Binding site: carbohydrate (Ser) (covalent) #status predicted  
 F:451,454,594/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 99.7%; Score 965; DB 2; Length 1315;  
 Best Local Similarity 99.5%; Pred. No. 4.8e-86;  
 Matches 183; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 HTHODFQPVHLVALNTPLSGKMGIRGADFOCFQOARAVGLSGTFRAPFLSSRLQDLXSI 60  
 |||||||  
 DB 1132 HTHODFQPVHLVALNTPLSGKMGIRGADFOCFQOARAVGLSGTFRAPFLSSRLQDLXSI 1191  
 |||||||

```

0Y      61 VRRARAGSVPIYNLKDEVLSPSMDSLFSGSGGQLDPPGARIFEDCDRDVLRHPAMPQKSYW  120
        |||||||
Db      1192 VRRARAGSVPIYNLKDEVLSPSMDSLFSGSGGQVPGGARIFEDCDRDVLRHPAMPQKSYW  1251

0Y      121 HGSDDSGRRMLMESYCEWTETTTGATGCASSLLSGRLLDQKKAASCHNSYIVLCIENSFMT  180
        |||||||
Db      1252 HGSDDSGRRMLMESYCEWTETTTGATGCASSLLSGRLLDQKKAASCHNSYIVLCIENSFMT  1311

0Y      181 SFSK 184
        |||||
Db      1312 SFSK 1315

RESULT      3
A:33019
collagen alpha 1(XVIII) chain - human (fragment)
M:Contains: endostatin
C:Species: Homo sapiens (man)
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 31-Mar-2000
C:Accession: A53019
R:Oh, S.-P.; Matman, M.L.; Seidlin, M.F.; Cheng, S.D.; Knoll, J.H.M.; Timmons, S.; Olsen,
Genomics 19, 494-499, 1994
A:Title: Cloning of cDNA and genomic DNA encoding human type XVIII collagen and localized
A:Reference numbers: A53019; MUID:94245237
A:Accession: A53019
A:Molecule type: mRNA
A:Residues: 1-684 <OH>
A:Cross-references: GB:122548; NID:9348908; PIDN:AA51864.1; PID:9562794
A:Note: The cited accession number, 125548, is not in Genbank release 103
A:Note: In the authors' translation, 482-gly is not shown, residues 483-490 are shifted
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
C:Comment: Different splice forms of collagen alpha 1(XVIII) may be involved in perivascu-
C:Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of un-
ay be useful in treating solid tumors.
C:Genetics:
A:Gene: GDB:COL18A1
A:Cross-references: GDB:138752; OMIM:120328
A:Map position: 21q22.3-21q22.3
C:Superfamily: unassigned collagens
C:Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglycan
F:1-684/Product: collagen alpha 1(XVIII) chain (fragment) #status predicted <MAT>
F:1-59/Domain: collagenous (fragment) #status predicted <CO4>
F:74-115/Domain: collagenous #status predicted <CO5>
F:129-201/Domain: collagenous #status predicted <CO6>
F:212-244/Domain: collagenous #status predicted <CO7>
F:257-278/Domain: collagenous #status predicted <CO8>
F:262-264/Region: cell attachment (R-G-D) motif
F:286-340/Domain: collagenous #status predicted <CO9>
F:354-371/Domain: collagenous #status predicted <CO10>
F:502-684/Product: endostatin #status predicted <EST>
F:509-684/Region: multiplexin collagen carboxyl-terminal similarity

Query Match      86.8%  Score 840;  DB 2;  Length 684;
Best Local Similarity 85.6%  Pred. No. 3.6e-74;
Matches 155;  Conservative 15;  Mismatches 11;  Indels 0;  Gaps 0;

0Y      1 HTHQDFQVLAHLVALNPLSLSGMRCIRGADFCQFOARAVGLSGTFRFLSLRLDLYSI  60
        |||||||
Db      502 HSHRDFQVLAHLVALNPLSLSGMRCIRGADFCQFOARAVGLAGTFRFLSLRLDLYSI  561

0Y      61 VRRADRGVPIYNLKDEVLSPSMDSLFSGSGGQLDPPGARITSPQCDRVLRHPAMPQKSYW  120
        |||||||
Db      562 VRRADRAVPIYNLKDELPLPSMEALFSGSGGQVPGGARITSPQCDRVLRHPAMPQKSYW  621

0Y      121 HGSDDSGRRMLMESYCEWTETTTGATGCASSLLSGRLLDQKKAASCHNSYIVLCIENSFMT  180
        |||||||
Db      622 HGSDDSGRRMLMESYCEWTETTTGATGCASSLLSGRLLDQKKAASCHNSYIVLCIENSFMT  681

0Y      181 S 181
        |||
Db      682 A 682

```

```

RESULT      4
A53317
collagen alpha 1(XV) chain precursor - human
M:Alternate names: procollagen alpha 1(XV) chain
C:Species: Homo sapiens (man)
G:Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 31-Mar-2000
J:Akiyama, K.; Miyazawa, T.; Nishida, Y.; Imai, S.; Tanabe, T.
R:Kiyakura, S.; Heinemann, P.; Rehm, M.; Honkanen, N.; Myers, J.C.; Phlatjanemi, R. Biol. Chem. 269, 4773-4779, 1994
A>Title: Primary structure of the alpha chain of human type XV collagen and exon-intron A:Reference number: A53317; MUID:94148920
A:Accession: A53317
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1388 <RIV>
A:CROSS-references: GB:I25280
A>Note: nucleotide sequence and conceptual translation not complete
R:Mutagaki, Y., Abe, N., Niimiya, Y., Olsen, B.R., Ooshima, A.
J. Biol. Chem. 269, 4042-4046, 1994
A>Title: The human alpha1(xv) collagen chain contains a large amino-terminal non-trip A:Reference number: A53146; MUID:94140817
A:Accession: A53146
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-9,'S','I'-48,'V','-50-94,'A','B'-149,'A','L51-203,'V','P'-205-408,'A','410-569 <MA
X>CROSS-references: GB:D1230; NIH:g415605; PIDN:BAAO4762.1; PID:d1005294; PID:g4607070
R:Myers, J.C.; Kiyakura, S.; Gordon, M.K.; Dion, A.S.; Phlatjanemi, T.
Proc. Natl. Acad. Sci. U.S.A. 89, 10144-10148, 1992
A>Title: Identification of a previously unknown human collagen chain, alpha1(XV), cha A:Reference number: S28778; MUID:93066196
A:Accession: S28778
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 544-640,'P','F',642-811,'P','E',813-1252 <MEY>
C:Genetics:
A:Gene: GDB:COL15A1
A:CROSS-references: GDB:132578; OMIM:120325
A:Map position: 9q21-q922
C:Superfamily: unassigned collagens
F:1-22/Domain: signal sequence #status predicted <SIG>
F:21-1388/Product: collagen alpha 1(XV) chain #status predicted <MAT>
F:1216-1388/Region: multiplexin collagen carboxyl-terminal similarity

Query Match          58.2%; Score 563; DB 2; Length 1388;
Best Local Similarity 60.7%; Pred. No. 1,1e+46;
Matches 108; Conservative 23; Mismatches 43; Indels 4; Gaps 1;

OY    7 OPVLALVALNTPLSSGMGIRGADPOCFQOKARAVGLSTFAFLSSRLQDLYSTRADR   66
       :|::|||:|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    1215 KPALLHALALNMFEFGSDIR---ADFQCFKAKRAAGLLSYRAFSLSHLDISTVRKAER 1270
               |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY    67 GSVPILVNKLDEVLSPMSDLFSFGSQGLQPAPRIFFSGFRGVLRHPAMPVKSVHGSPPS 126
       ||::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    1271 YSLPIVLNKGYLVFNMMQSIFSGHGCGFMNHIPITSEFCGRIMDPSPQPVVIHHGSSPH 1330
               ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY    127 GRRLMESICEFTWRPFETTGATGOASSLSGLRLFLEOKAASCHNSYLVLCIENTSFMTSSF 184
       ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    1331 GVRLVDNYCEAMWRTADTVAVTGTLASPLSTGKITLDOKAYSCANRLIVLICENSMFDARK 1388
               ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT      5
T22002
hypothetical protein F39H11.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22002
R:white, S.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19500
A:Accession: T22002
```



A>Title: The agrin gene codes for a family of basal lamina proteins that differ in function  
 A:Reference number: A38857; MUID:92232298  
 A:Contents: alternative splicing  
 A:Accession: A38857  
 A:Molecule type: mRNA  
 A:Residues: 1132-1783; 1795-1955 <R02>  
 A:Cross-references: GB:M97371  
 A:Accession: B38857  
 A:Molecule type: mRNA  
 A:Residues: 1221-1647; 1652-1783; 1794-1955 <R03>  
 A:Cross-references: GB:M97372  
 A>Note: translation of the nucleotide sequence is not complete  
 R:Thomas, W.S.; O'Dowd, D.K.; Smith, M.A.  
 Dev. Biol. 158, 523-535, 1993  
 A>Title: Developmental expression and alternative splicing of chick agrin RNA.  
 A:Reference number: 150692; MUID:93345745  
 A:Accession: 150692  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 'SHLSNEIPA', 1784-1795 <THO>  
 A:Cross-references: EMBL:U07271; NID:9459665; PIDN:AA16788.1; PID:9459666  
 C:Comment: This protein mediates the motor neuron-induced aggregation of acetylcholine A  
 C:Comment: Alternative splicing produces two inactive proteins: agrin-related protein 1  
 C:Keywords: alternative splicing; duplication; glycoprotein; neuromuscular junction  
 F:1-38/Domain: signal sequence #status predicted <SIG>  
 F:39-1783/Domain: agrin #status predicted <MAT>  
 F:39-1783, 1795-1955/Product: agrin-related protein 1 #status predicted <AG1>  
 F:39-1647, 1652-1783, 1794-1955/Product: agrin-related protein 2 #status predicted <AG2>  
 F:77-126/Domain: kazal proteinase inhibitor homology <KPI1>  
 F:152-201/Domain: kazal proteinase inhibitor homology <KPI2>  
 F:223-273/Domain: kazal proteinase inhibitor homology <KPI3>  
 F:293-344/Domain: kazal proteinase inhibitor homology <KPI4>  
 F:379-418/Domain: kazal proteinase inhibitor homology <KPI5>  
 F:435-483/Domain: kazal proteinase inhibitor homology <KPI6>  
 F:500-548/Domain: kazal proteinase inhibitor homology <KPI7>  
 F:584-633/Domain: kazal proteinase inhibitor homology <KPI8>  
 F:678-726/Domain: laminin-type EGF-like homology <LE1>  
 F:729-773/Domain: laminin-type EGF-like homology <LE2>  
 F:801-851/Domain: kazal proteinase inhibitor homology <KPI9>  
 F:856-995/Region: serine/threonine-rich  
 F:1150-1219/Region: serine/threonine-rich  
 F:1233-1264/Domain: BGF homology <EG1>  
 F:1294-1448/Domain: laminin G repeat homology <LG1>  
 F:1429-1431/Region: motor neuron attachment (L-R-E) motif  
 F:1450-1482/Domain: EGF homology <EG2>  
 F:1489-1521/Domain: BGF homology <EG3>  
 F:1560-1711/Domain: laminin G repeat homology <LG2>  
 F:1718-1751/Domain: BGF homology <EG4>  
 F:1803-1955/Domain: laminin G repeat homology <LG3>  
 F:186-105, 94-126, 160-180, 169-201, 233-252, 241-273, 304-323, 312-344, 378-397, 386-418, 443-462, 1482, 1489-1500, 1494-1510, 1512-1521/Disulfide bonds: #status predicted  
 F:390, 659, 764, 814/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 8.1%; Score 78; DB 1; Length 1955;  
 Best Local Similarity 22.3%; Pred. No. 44;  
 Matches 47; Conservative 24; Mismatches 62; Indels 78; Gaps 9;

OY 5 DEQPLHLVAVLTPISGGR--GIRGADQCQOARAVGLSTFRF----- 49  
 DB 1670 DESKLARAALSTSTFYGAQORISIKGVPLKEQHRSAVESITFRANPCTOKRPNQONGC 1729  
 OY 50 -LSSRLQDLYSIVRRADSGSVIVMLKDEVLSPSWDSLFGSGQ-----GQLQPGARIFS 102  
 DB 1730 TCSPLRESEACQGR-----FSGHCKEVIITEKAAGDAEALA 1767  
 OY 103 PDGR-----DYLRRHPAMPQKSVWHGSDPSGRRIMESYCE-TWRTETTGATGQ 148  
 DB 1768 FDGRTYMEYHNAVTKSPALDPA-----EPSEKALQSNHPELSIKREAT-----Q 1813  
 OY 149 ASLSLGRLEOKAASCHNSYIVLCIENFSM 179

DB 1814 GLILWSGKGLR-----SDYIALAIVDGFV 1838  
 RESULT 9  
 T10011  
 probable penicillin-binding protein - Mycobacterium leprae  
 C:Species: Mycobacterium leprae  
 C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 16-Jul-1999  
 C:Accession: T10011  
 R:Cole, S.T.  
 submitted to the EMBL Data Library, August 1997  
 A:Reference number: Z16916  
 A:Accession: T10011  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-474 <COL>  
 A:Cross-references: EMBL:Z70722; NID:e1059634; PID:e338505  
 C:Genetics:  
 A:Gene: pbpa

Query Match 8.0%; Score 77.5; DB 2; Length 474;  
 Best Local Similarity 24.0%; Pred. No. 8.8;  
 Matches 35; Conservative 18; Mismatches 68; Indels 25; Gaps 5;

OY 35 OQARVGLSGTFRALFSLRQDLSTVRADSGSVIVMLKDEVLSPSWDSLFGSGQGL 94  
 DB 101 EDAEDALWGSDBRLFGRLADFT--GDPDGGAVDTTINRVQOTGMDAQGGGSGP 158  
 OY 95 QPCARIF-\*\*\*\*\*SFDGRDYLRRHPAMPQKSVWH--GSDPS-----GRRIMES 133  
 DB 159 CGAVVALEPSYNGKILAWSTPSYDPLNLASHNPEQAQAMRLDHPSPILNRAISFT 218  
 OY 134 Y--CEWRRETTGATGQASLSLGR 157  
 DB 219 YPPGSTFKVITTTAALQAGATTSDOL 244

RESULT 10  
 T30283  
 polyketide synthase - Streptomyces sp. (strain MA6548)  
 C:Species: Streptomyces sp.  
 A:Variety: strain MA6548  
 C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 17-Nov-2000  
 C:Accession: T30283  
 R:Motamedi, H.; Cai, S.J.; Shafiee, S.J.; Elliston, K.O.  
 Eur. J. Biochem. 244, 74-80, 1997  
 A>Title: Structural organization of a multifunctional polyketide synthase involved in  
 A:Reference number: Z20806; MUID:97217427  
 A:Accession: T30283  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-6420 <MOT>  
 A:Cross-references: EMBL:Y10438; NID:e1014806; PID:e290681; PIDN:CAA71463.1  
 C:Genetics:  
 A:Note: fkbA  
 C:Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier pro  
 C:Keywords: carrier protein  
 F:51-433/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>  
 F:1930-2325/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>  
 F:1331-3462/Domain: acyl carrier protein homology <ACP1>  
 F:3505-3900/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>  
 F:3983-4254/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>  
 F:5307-5378/Domain: acyl carrier protein homology <ACP2>  
 F:5431-5831/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>  
 F:5932-6206/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>  
 F:6293-6364/Domain: acyl carrier protein homology <ACP3>

Query Match 8.0%; Score 77.5; DB 2; Length 6420;  
 Best Local Similarity 23.0%; Pred. No. 2.1e+02;  
 Matches 47; Conservative 25; Mismatches 73; Indels 59; Gaps 9;

```

0Y      , 4 ODFGVHLHVALNPLSCGMR-----GIRGADQCFQO-----A 37
Db      5406 RDQDEPILAIYMAACLPQGVISPERLMRLVESGTDALITTAFGDRGMDLDALITDPPDAVG 5465
0Y      38 RAVGLSGTFRAFLSSRLDLYSIYBRADGSPYVNLKDEVLSPSMDSLFGSGQGLQPG 97
Db      5466 KAYSLRGFLLEGAAFEFDAAFDISPRESLQMDPQRL---LLETWMEAI-----ERGRINPA 5519
0Y      98 ARIISFQDGRDVLIRHPAMPQKSVYHSGSDPSGRRLMESYCEWTFETTTGATGQANSSLLSRL 157
Db      5520 ----SIHQREI-----GVYGAAGQGYGLGAE-----DTEGNATITGSGTILLSRL 5561
0Y      158 -----LEOKA-----ASCHNSYIVL 172
Db      5562 AVYLGLEGPAYVTVDTCASSLVAL 5585

```







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2001, 12:50:51 ; Search time 11.67 Seconds

(without alignments)  
540.104 Million cell updates/sec

Title: US-09-589-777A-2

Perfect score: 968  
Sequence: 1 HTHODFQPLHLVLTPLS.....CHNSYLVLCIENSEMFSFK 184

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	968	100.0	1527	1 CAIH_MOUSE	P39061 mus musculu
2	840	86.8	1516	1 CAIH_HUMAN	P39060 homo sapien
3	563	58.2	1388	1 CAIE_HUMAN	P39059 homo sapien
4	78	8.1	1955	1 AGRI_CHICK	P31696 gallus gall
5	77	8.0	1574	1 SYUL_RAT	062910 rattus norv
6	76.5	7.9	1050	1 ULK1_HUMAN	075385 homo sapien
7	76	7.9	309	1 YHCC_ECOLI	P45476 escherichia
8	76	7.9	1575	1 SYU1_HUMAN	043426 homo sapien
9	75	7.7	505	1 CATI_METBA	03662 methanosarc
10	75	7.7	587	1 HEMO_HUMAN	P22557 homo sapien
11	75	7.7	1324	1 SYUL_BOVIN	018964 bos tauru
12	74.5	7.7	611	1 PHBC_RHIME	P50176 r poly-beta
13	74.5	7.6	512	1 PPX_ECOLI	P29014 pseudomonas
14	72.5	7.5	419	1 P47K_PSECL	P31521 pseudomonas
15	72.5	7.5	739	1 VNUC_EBOGA	Q9qce9 ebola virus
16	72.5	7.5	800	1 P5CS_GAEEL	P64889 c probable
17	72	7.4	1233	1 NME3_HUMAN	014957 homo sapien
18	71.5	7.4	692	1 GYRB_BARBA	P94281 bartonella
19	71.5	7.4	5217	1 HTSL_COCCA	Q01886 cochlidiolu
20	71	7.3	587	1 HEMO_RAT	063147 rattus norv
21	71	7.3	1289	1 CSAB_BACUD	Q45753 bacillus th
22	70.5	7.3	257	1 FAEL_ECOLI	P33784 escherichia
23	70.5	7.3	316	1 YEEY_ECOLI	P76369 escherichia
24	70.5	7.3	319	1 HALL_RAT	P15978 rattus norv
25	69.5	7.2	437	1 ADFF_HUMAN	Q99541 homo sapien
26	69.5	7.2	738	1 VNUC_EBOSB	Q9qj77 ebola virus
27	69.5	7.2	739	1 VNUC_EBOZ5	072142 ebola virus
28	69.5	7.2	739	1 VNUC_EBOZ5	072142 ebola virus
29	69.5	7.2	860	1 SZLA_BRARE	09w711 brachydanio
30	69	7.1	174	1 CARO_MYXA	Q06909 myxococcus
31	69	7.1	692	1 VNUC_MABVM	P27588 marburg vir
32	69	7.1	695	1 VNUC_MABVM	P25263 marburg vir
33	69	7.1	1023	1 DPOL_ADEB3	072540 bovine aden

## ALIGNMENTS

RESULT	ID	CAIH_MOUSE	STANDARD	PRT	1527 AA
34	69	7.1	1357	1	VY03_YEAST
35	68.5	7.1	266	1	NANH_BACR
36	68.5	7.1	449	1	PURL_PYRO
37	68.5	7.1	1248	1	STJ2_RAT
38	68	7.0	362	1	SERC_BACCI
39	68	7.0	679	1	RNSA_MOUSE
40	68	7.0	1051	1	ULK1_MOUSE
41	68	7.0	1203	1	PRK2_YEAST
42	68	7.0	1220	1	CSAC_BACTU
43	68	7.0	1385	1	CSAA_BACUD
44	68	7.0	1443	1	SYJ2_HUMAN
45	67.5	7.0	230	1	VHEL_P1AMV
P47104 saccharomyc					
P31206 bacteroides					
O57979 pyrococcus					
O55207 rattus norv					
O59196 bacillus th					
O05921 mus musculu					
O70405 mus musculu					
P22138 saccharomyc					
P36935 bacillus th					
O45760 bacillus th					
O15056 homo sapien					
O07517 plantago as					
RESULT 1					
CAIH_MOUSE					
AC	P39061	062002	061437		
DT	01-FEB-1995	(Rel. 31, Created)			
DT	01-OCT-2000	(Rel. 40, Last sequence update)			
DT	01-OCT-2000	(Rel. 40, Last annotation update)			
DE	COLLAGEN ALPHA 1(XVII) CHAIN PRECURSOR [CONTAINS: ENDOSTATIN].				
GN	COL18A1.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
OX	NCBI_Taxid:10090;				
RN	[1]				
RP	SEQUENCE FROM N.A. (SHORT ISOFORM).				
RC	STRAIN-BALB/C; TISSUE-Liver;				
RX	MEDLINE-94245707; PubMed-818673;				
RA	Rehn M.V., Hintikka E., Philajaniemi T.;				
RT	"Primary structure of the alpha 1 chain of mouse type XVII collagen,				
RT	partial structure of the corresponding gene, and comparison of the				
RT	alpha 1(XVII) chain with its homologue, the alpha 1(XV) collagen				
RL	J. Biol. Chem. 269:13929-13935(1994).				
RN	[2]				
RP	SEQUENCE FROM N.A. (SHORT AND LONG ISOFORMS).				
RA	Rehn M., Hintikka E., Philajaniemi T.;				
RT	"Characterization of the mouse gene for the alpha-1 chain of type				
RT	XVIII collagen (COL18A1) reveals that the three variant N-terminal				
RT	polypeptide forms are transcribed from two widely separated				
RT	promoters.";				
RL	submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE OF 213-1140 FROM N.A. (SHORT ISOFORM).				
RX	MEDLINE-94240112; PubMed-818394;				
RA	Rehn M.V., Philajaniemi T.;				
RT	"Alpha 1(XVIII), a collagen chain with frequent interruptions in the				
RT	collagenous sequence, a distinct tissue distribution, and homology				
RT	with type XV collagen.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).				
RN	[4]				
RP	SEQUENCE OF 240-1527 FROM N.A.				
RC	TISSUE-Liver;				
RX	MEDLINE-94240111; PubMed-818393;				
RA	Oh S.P., Kamagata Y., Timmons S., Ooshima A., Olsen B.R.;				
RT	"Isolation and sequencing of cDNAs for proteins with multiple domains				
RT	of Gly-xaa-yaa repeats identify a distinct family of collagenous				
RT	proteins.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 91:4229-4233(1994).				
RN	[5]				
RP	CHARACTERIZATION OF ENDOSTATIN, AND PARTIAL SEQUENCE.				
RX	MEDLINE-97160848; PubMed-9008168;				
RA	O'Reilly M.S., Boehm T., Shing Y., Fukai N., Vastios G., Lane W.S.,				
RA	Flynn E., Birkhead J.R., Olsen B.R., Folkman J.;				
RT	"Endostatin: an endogenous inhibitor of angiogenesis and tumor				
RT	growth.";				
RL	Cell 88:277-285(1997).				

[6]  
 RN X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF ENDOSTATIN.  
 RP MEDLINE-98169382; PubMed-9501087;  
 RA Hohenester E., Sasaki T., Olsen B.R., Timpl R.;  
 RT "Crystal structure of the angiogenesis inhibitor endostatin at 1.5-A  
 resolution.";  
 RL EMBO J. 17:1656-1664(1998).  
 CC -1- FUNCTION: ENDOSTATIN POTENTIALLY INHIBITS ENDOTHELIAL CELL  
 CC PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY  
 CC BINDING TO THE HEPARAN SULPHATE PROTEOGLYCANS INVOLVED IN GROWTH  
 CC FACTOR SIGNALING.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A SHORT FORM AND A LONG FORM  
 CC (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- SIMILARITY: BELONGS TO THE MULTIPLEXIN FAMILY OF COLLAGENS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: I16898; AAA37434.1; -  
 DR EMBL: U03714; AAA20657.1; -  
 DR EMBL: U03715; AAC52901.1; -  
 DR EMBL: U34606; AAC52901.1; JOINED.  
 DR EMBL: U34608; AAC52901.1; JOINED.  
 DR EMBL: U34609; AAC52901.1; JOINED.  
 DR EMBL: U34610; AAC52901.1; JOINED.  
 DR EMBL: U34611; AAC52901.1; JOINED.  
 DR EMBL: U34612; AAC52901.1; JOINED.  
 DR EMBL: U34613; AAC52901.1; JOINED.  
 DR EMBL: U03716; AAC52901.1; JOINED.  
 DR EMBL: U03717; AAC52901.1; JOINED.  
 DR EMBL: U03718; AAC52902.1; -  
 DR EMBL: U34607; AAC52902.1; JOINED.  
 DR EMBL: U34608; AAC52902.1; JOINED.  
 DR EMBL: U34609; AAC52902.1; JOINED.  
 DR EMBL: U34610; AAC52902.1; JOINED.  
 DR EMBL: U34611; AAC52902.1; JOINED.  
 DR EMBL: U34612; AAC52902.1; JOINED.  
 DR EMBL: U34613; AAC52902.1; JOINED.  
 DR EMBL: U03716; AAC52902.1; JOINED.  
 DR EMBL: U03718; AAC52902.1; JOINED.  
 DR EMBL: U11636; AAC52178.1; -  
 DR EMBL: I22545; AAA19787.1; -  
 DR PDB: 1KOE; 16-FEB-99.  
 DR MGI: 88451; COL18a1.  
 DR InterPro: IPR000087; -  
 DR Pfam: PF01391; Collagen; 6.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;  
 KW 3D-structure.  
 FT CHAIN 1 26 POTENTIAL.  
 FT CHAIN 27 1527 COLLAGEN ALPHA 1(XVIII) CHAIN.  
 FT CHAIN 1344 1527 ENDOSTATIN.  
 FT DOMAIN 27 538 NONHELICAL REGION 1 (NC1).  
 FT DOMAIN 539 555 TRIPLE-HELICAL REGION 1 (COL1).  
 FT DOMAIN 556 575 NONHELICAL REGION 2 (NC2).  
 FT DOMAIN 576 649 TRIPLE-HELICAL REGION 2 (COL2).  
 FT DOMAIN 650 673 NONHELICAL REGION 3 (NC3).  
 FT DOMAIN 674 795 TRIPLE-HELICAL REGION 3 (COL3).  
 FT DOMAIN 796 818 NONHELICAL REGION 4 (NC4).  
 FT DOMAIN 819 901 TRIPLE-HELICAL REGION 4 (COL4).  
 FT DOMAIN 902 915 NONHELICAL REGION 5 (NC5).  
 FT DOMAIN 916 957 TRIPLE-HELICAL REGION 5 (COL5).  
 FT DOMAIN 958 970 NONHELICAL REGION 6 (NC6).  
 FT DOMAIN 971 1043 TRIPLE-HELICAL REGION 6 (COL6).  
 FT DOMAIN 1044 1053 NONHELICAL REGION 7 (NC7).  
 FT DOMAIN 1054 1086 TRIPLE-HELICAL REGION 7 (COL7).

FT	DOMAIN	1087	1098	NONHELICAL REGION 8 (NC8).
FT	DOMAIN	1099	1122	TRIPLE-HELICAL REGION 8 (COL8).
FT	DOMAIN	1123	1129	NONHELICAL REGION 9 (NC9).
FT	DOMAIN	1130	1181	TRIPLE-HELICAL REGION 9 (COL9).
FT	DOMAIN	1182	1194	NONHELICAL REGION 10 (NC10).
FT	DOMAIN	1195	1212	TRIPLE-HELICAL REGION 10 (COL10).
FT	DOMAIN	1213	1527	NONHELICAL REGION 11 (NC11).
FT	CARBOHYD	338	338	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	700	700	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	DISULFID	1376	1516	
FT	DISULFID	1478	1508	
FT	SITE	1104	1106	
FT	VARSPLIC	1	212	
FT	VARSPLIC	213	238	
FT	CONFLICT	900	900	CELL ATTACHMENT SITE (POTENTIAL).
FT	CONFLICT	947	947	MISSING (IN SHORT ISOFORM).
FT	CONFLICT	964	964	APPTLPPEOSNLAIPRSPAPDF -> MAPRHLIDVL
FT	CONFLICT	1157	1157	TSVLTLVAVSMAL (IN SHORT ISOFORM).
FT	CONFLICT	1266	1266	
FT	CONFLICT	1276	1276	
FT	CONFLICT	1437	1437	
FT	CONFLICT	1527	1527	
FT	SEQUENCE	1527	15608	MM; 9645045AF1408513 CRC64;

Query Match 100.0%; Score 968; DB 1; Length 1527;  
 Best Local Similarity 100.0%; Pred. No. 3,2e-84;  
 Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	HHHODEPVLHLVNLPTLSGGNRRIGADFOCFQDARAAGLSGTFRAFLSRRLDLYSI	60
DB	1344	HHHODEPVLHLVNLPTLSGGNRRIGADFOCFQDARAAGLSGTFRAFLSRRLDLYSI	1403
QY	61	VRRADRGSPYVNLKDEVLSFSSGQGLQDQGARIFSPDGDVLRHPAMPQKSW	120
DB	1404	VRRADRGSPYVNLKDEVLSFSSGQGLQDQGARIFSPDGDVLRHPAMPQKSW	1463
QY	121	HGSDPSGRRLMESYCEMTTETTGATGQASSLSGRHLQKKAASCHNSYIVLCIENSFMT	180
DB	1464	HGSDPSGRRLMESYCEMTTETTGATGQASSLSGRHLQKKAASCHNSYIVLCIENSFMT	1523
QY	181	SPSK 184	
DB	1524	SPSK 1527	

RESULT 2  
 CA1H\_HUMAN STANDARD; PRT; 1516 AA.  
 AC P39060; Q9Y608; Q9Y607; Q9UK38;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE COLLAGEN ALPHA 1(XVIII) CHAIN PRECURSOR [CONTAINS: ENDOSTATIN].  
 GN COL18A1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euteria; Primates; Catarrhini; Hominoidea; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-98164096; PubMed-9503365;  
 RA Saarela J., Ylikarppa R., Rehn M., Purmonen S., Pihlajaniemi T.;  
 RT "Complete primary structure of two variant forms of human type XVIII  
 RT collagen and tissue-specific differences in the expression of the  
 RT corresponding transcripts.";  
 RL Matrix Biol. 16:319-328(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-20289799; PubMed-10830953;  
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,  
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,  
 RA Ohki M., Takagi T., Sasaki Y., Taudien S., Blechschmidt K., Polley A.,

RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,  
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,  
RA Rosenthal A., Kuch J., Shibuya K., Kawasaki K., Asakawa S.,  
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,  
RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,  
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,  
RA Kanser J., Beck A., Klages S., Hennig S., Rteselmann L., Dagand E.,  
RA Weinmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,  
RA Lehnach H., Reinhardt R., Yaspo M.-L.,  
RA "The DNA sequence of human chromosome 21." Nature 405:311-319(2000).  
RL Nature 405:311-319(2000).  
RP [3] SEQUENCE OF 834-1516 FROM N.A.  
RP MEDLINE=94245237; PubMed=8188291;  
RX ON S.P., Worman M.L., Seldin M.F., Cheng S., Knoll J.H., Timmons S.,  
RT Olsen B.R.;  
RT "Cloning of cDNA and genomic DNA encoding human type XVIII collagen  
RT and localization of the alpha 1(XVIII) collagen gene to mouse  
RT chromosome 10 and human chromosome 21." Genomics 19:494-499(1994).  
RL [4] SEQUENCE OF 1314-1516 FROM N.A.  
RP TISSUE-Placenta:  
RC Zhi-Yong H., Biao L., Wei-Jie Z., Xiang-Fu W.;  
RT "Cloning and expression of human endostatin gene in Escherichia  
RT coli." Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
RL [5] INVOLVEMENT IN KNOBLOCH SYNDROME.  
RP MEDLINE=20400145; PubMed=10942434;  
RX Sertle A.L., Sossi V., Camargo A.A., Zatz M., Brähe C.,  
RA Passos-Bueno M.R.;  
RT "Collagen XVIII, containing an endogenous inhibitor of angiogenesis  
RT and tumor growth, plays a critical role in the maintenance of retinal  
RT structure and in neural tube closure." Hum. Mol. Genet. 9:2051-2058(2000).  
RL [6] FUNCTION: COL1A1A PROBABLY PLAYS A MAJOR ROLE IN DETERMINING THE  
CC RETINAL STRUCTURE AS WELL AS IN THE CLOSURE OF THE NEURAL TUBE.  
CC [7] FUNCTION: ENDOSTATIN POTENTIALLY INHIBITS ENDOTHELIAL CELL  
CC PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY  
CC BINDING TO THE HEPARAN SULFATE PROTEOGLYCANS INVOLVED IN GROWTH  
CC FACTOR SIGNALING (BY SIMILARITY).  
CC [8] ALTERNATIVE PRODUCTS: 2 ISOFORMS; A SHORT FORM/NC1-303 AND A LONG  
CC FORM/NC-493 (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC [9] TISSUE SPECIFICITY: PRESENT IN MULTIPLE ORGANS WITH HIGHEST LEVELS  
CC IN LIVER, LUNG AND KIDNEY.  
CC [10] PM: PROLINS AT THE THIRD POSITION OF THE TRIPETIDE REPEATING  
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
CC [11] DISEASE: DEFECTS IN COL1A1 ARE A CAUSE OF KNOBLOCH SYNDROME (KS);  
CC AN AUTOSOMAL RECESSIVE DISORDER DEFINED BY THE OCCURRENCE OF HIGH  
CC MYOPIA, VITREORETINAL DEGENERATION WITH RETINAL DETACHMENT,  
CC MACULAR ABNORMALITIES AND OCCIPITAL ENCEPHALOCELE.  
CC [12] SIMILARITY: BELONGS TO THE MULTIPLEXIN FAMILY OF COLLAGENS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL database -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as their content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

KW Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing.  
FT SIGNAL 1 23  
FT CHAIN 24 1516  
FT CHAIN 1334 1516  
FT CHAIN 24 516  
FT DOMAIN 24 516  
FT DOMAIN 517 550  
FT DOMAIN 551 560  
FT DOMAIN 561 640  
FT DOMAIN 641 664  
FT DOMAIN 665 786  
FT DOMAIN 787 809  
FT DOMAIN 810 892  
FT DOMAIN 893 906  
FT DOMAIN 907 948  
FT DOMAIN 949 961  
FT DOMAIN 962 1034  
FT DOMAIN 1035 1044  
FT DOMAIN 1045 1077  
FT DOMAIN 1078 1089  
FT DOMAIN 1090 1111  
FT DOMAIN 1112 1118  
FT DOMAIN 1119 1173  
FT DOMAIN 1174 1186  
FT DOMAIN 1187 1204  
FT DOMAIN 1205 1516  
FT CARBOHYD 68 68  
FT CARBOHYD 129 129  
FT CARBOHYD 164 164  
FT CARBOHYD 691 691  
FT DISULFID 1366 1506  
FT SITE 1498 1498  
FT SITE 1095 1097  
FT SITE 1 180  
FT VARSPLIC 181 215  
FT VARSPLIC 428 428  
FT CONFLICT 841 877  
FT CONFLICT 886 886  
FT CONFLICT 912 912  
FT CONFLICT 933 933  
FT CONFLICT 975 975  
FT CONFLICT 1064 1064  
FT CONFLICT 1084 1084  
FT CONFLICT 1120 1120  
FT CONFLICT 1123 1123  
FT CONFLICT 1126 1126  
FT CONFLICT 1206 1206  
FT CONFLICT 1304 1304  
FT CONFLICT 1314 1314  
FT CONFLICT 1323 1324  
FT CONFLICT 1437 1437  
FT CONFLICT 1443 1443  
FT CONFLICT 1483 1483  
FT SEQUENCE 1516 AA; 153840 MM; 3C70F29AA4476EB76 CRC64;  
Query Match 86.8%; Score 840; DB 1; Length 1516;  
Best Local Similarity 85.6%; Pred. No. 4.9e-72;  
Matches 155; Conservative 15; Mismatches 11; Indels 0; Gaps 0;  
OY 1 HTTODPQVYLVLALNPLSGMGIGICGAPGOCQOARAVGLSTPFAFLSSRLQDLYST 60  
DB 1334 RSHDDPQVYLVLALNPLSGMGIRGADQOQOARAVGLATFAFLSSRLQDLYST 1393  
OY 61 VRRADRGVPIVNLKDEVLSPSMDLSFGSGQLOPQARIFSFGRVYLHHPAMPQKSVW 120  
DB 1394 VRRADRAAVPIVNLKDEVLSPSMDLSFGSGQLOPQARIFSFGRVYLHHPAMPQKSVW 1453  
OY 121 HGSDPSRRRLTESYCEFTWRTEGATQOASSLLGRLLEKKAASCHNSYVLICENSFMT 180  
DB 1454 HGSDPNGRRRLTESYCEFTWRTEGATQOASSLLGRLLEKKAASCHNSYVLICENSFMT 1513

QY 181 S 181  
DB 1514 A 1514

RESULT 3  
CAIE\_HUMAN STANDARD: PRT: 1388 AA.  
ID CAIE\_HUMAN  
AC P39059;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE COLLAGEN ALPHA 1(XV) CHAIN PRECURSOR.  
GN COL15A1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=umbilical cord;  
RX MEDLINE=94148920; PubMed=8106446;  
RA Kivirikko S., Heineken P., Rehn M.V., Honkanen N., Myers J.C.,  
RA Pihlajaniemi T.;  
RT "Primary structure of the alpha 1 chain of human type XV collagen and  
RT exon-intron organization in the 3' region of the corresponding  
RT gene";  
RT J. Biol. Chem. 269:4773-4779(1994).  
RN [2]  
RP SEQUENCE OF 1-569 FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=94140817; PubMed=8307960;  
RA Murgaki Y., Abe N., Ninomiya Y., Olsen B.R., Ooshima A.;  
RT "The human alpha 1(XV) collagen chain contains a large amino-terminal  
RT non-triple helical domain with a tandem repeat structure and homology  
RT to alpha 1(XVIII) collagen";  
RT J. Biol. Chem. 269:4042-4046(1994).  
RN [3]  
RP SEQUENCE OF 544-1252 FROM N.A.  
RX MEDLINE=93066196; PubMed=1279671;  
RA Myers J.C., Kivirikko S., Gordon M.K., Dion A.S., Pihlajaniemi T.;  
RT "Identification of a previously unknown human collagen chain, alpha  
RT 1(XV), characterized by extensive interruptions in the triple-helical  
RT region";  
RT Proc. Natl. Acad. Sci. U.S.A. 89:10144-10148(1992).  
RN [4]  
RP TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN INTERNAL ORGANS  
CC -1- TISSUE: ADRENAL GLAND, PANCREAS AND KIDNEY.  
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
CC -1- SIMILARITY: BELONGS TO THE MULTIPLEXIN FAMILY OF COLLAGENS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: I25286; AAAS8429.1; -  
CC EMBL: D21230; BAA04762.1; -  
CC EMBL: L01697; -; NOT\_ANNOTATED\_CDS.  
CC MIM: 120325; -  
DR InterPro: IPRO000087; -  
DR Pfam: PF01391; Collagen; 4.  
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Cell adhesion; Collagen; Glycoprotein; Signal.  
FT CHAIN 1 25  
FT SIGNAL 25  
FT CHAIN 26 1388  
FT DOMAIN 26 555  
FT DOMAIN 556 573  
FT DOMAIN 574 618  
NONHELIICAL REGION 1 (NC1).  
TRIPLE-HELIICAL REGION 1 (COL1).  
NONHELIICAL REGION 2 (NC2).

FT DOMAIN 619 732  
FT DOMAIN 733 763  
FT DOMAIN 764 798  
FT DOMAIN 799 822  
FT DOMAIN 823 867  
FT DOMAIN 868 878  
FT DOMAIN 879 949  
FT DOMAIN 950 983  
FT DOMAIN 984 1013  
FT DOMAIN 1014 1027  
FT DOMAIN 1028 1045  
FT DOMAIN 1046 1052  
FT DOMAIN 1053 1107  
FT DOMAIN 1108 1117  
FT DOMAIN 1118 1132  
FT DOMAIN 1133 1388  
FT DOMAIN 358 555  
FT DOMAIN 358 555  
FT REPEAT 408 408  
FT REPEAT 409 459  
FT REPEAT 460 509  
FT REPEAT 510 555  
FT REPEAT 510 555  
FT CARBOHYD 306 306  
FT CARBOHYD 324 324  
FT CARBOHYD 687 687  
FT CARBOHYD 807 807  
FT CARBOHYD 814 814  
FT CARBOHYD 1046 1046  
FT CONFLICT 10 10  
FT CONFLICT 49 49  
FT CONFLICT 95 95  
FT CONFLICT 150 150  
FT CONFLICT 204 204  
FT CONFLICT 409 409  
SQ SEQUENCE 1388 AA: 141930 MW: 60822AD925A3093D CRC64:

Query Match 58.2%; Score 563; DB 1; Length 1388;  
Best Local Similarity 60.7%; Pred. No. 1e-45;  
Matches 108; Conservative 23; Mismatches 43; Indels 4; Gaps 1;

QY 7 QYVLHVLAANTPLSGMRIRGADPQCFQARAVGLSGTFRFAFLSSRLDDLSIYRRDR 66  
DB 1215 KPLHLIALAMPFSGGIR--ADPQCFQARAAGLSYRAFSLSHDDLSIYRKAR 1270  
QY 67 GSVPIYNLKDEVLPSPMDLFGSGOGLDQGARIFSFDRDLRLHPAMPQKSVHMGSPDS 126  
DB 1271 YSLPIYNLKGVLFNNMDSIFSGHGQFMNHTPIYSFDRDLMDTPSPQKXVHMGSSPH 1330  
QY 127 GRLMESYCEWTETTGATGGAASSLSGRLLDQKRAASCHSYIVLCIENSFTSFSK 184  
DB 1331 GVALVONCEAMRTADTAVTGLASPLSTGKILDKAYSCANRLIVLCIENSFTMDARK 1388

RESULT 4  
AGRI\_CHICK STANDARD: PRT: 1955 AA.  
ID AGRI\_CHICK  
AC P31696;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE AGRIIN PRECURSOR.  
GN AGRIIN.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=92232297; PubMed=1314620;  
RA Tsim K.W.K., Ruegg M.A., Escher G., Kroeger S., McMahon U.J.;  
RT "cdna that encodes active agrin";

RL	Neuron 8:677-689(1992).	
RM	[2]	
RP	ALTERNATIVE SPLICING.	
RX	MEDLINE-92232298; PubMed-1314621;	
RA	Ruegg M.A., Tsim K.W.K., Horton S.E., Kroege S., Escher G.,	
RA	Gensch E.M., McMahon U.J.;	
RT	"The agrin gene codes for a family of basal lamina proteins that	
RT	differ in function and distribution.";	
RL	Neuron 8:691-699(1992).	
CC	-1- FUNCTION: COMPONENT OF THE BASAL LAMINA THAT CAUSES THE	
CC	AGGREGATION OF ACETYLCHOLINE RECEPTORS AND ACETYLCHOLINE-ESTERASE	
CC	ON THE SURFACE OF MUSCLE FIBERS OF THE NEUROMUSCULAR JUNCTION.	
CC	-1- SUBCELLULAR LOCATION: SYNAPTIC BASAL LAMINA AT THE NEUROMUSCULAR	
CC	JUNCTION.	
CC	-1- ALTERNATIVE PRODUCTS: AT LEAST THREE DIFFERENT FORMS ARISE BY	
CC	CLUSTERING ACTIVITY.	
CC	-1- SIMILARITY: CONTAINS 8 KAZAL-LIKE DOMAINS.	
CC	-1- SIMILARITY: CONTAINS 2 LAMININ EGF-LIKE DOMAINS.	
CC	-1- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> )	
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> )	
CC	-----	
DR	EMBL; M94271; AAA48585.1; -;	
DR	EMBL; M97371; AAA48586.1; -;	
DR	EMBL; M97372; -; NOT_ANNOTATED_CDS.	
DR	PIR; JH0591; AGCH.	
DR	InterPro: IPR000082; -	
DR	InterPro: IPR000152; -	
DR	InterPro: IPR000561; -	
DR	InterPro: IPR001239; -	
DR	InterPro: IPR00191; -	
DR	InterPro: IPR002049; -	
DR	InterPro: IPR002350; -	
DR	Pfam: PF00008; EGF; 4.	
DR	Pfam: PF01390; SEA; 1.	
DR	Pfam: PF00050; kazal; 9.	
DR	Pfam: PF00053; laminin_EGF; 2.	
DR	Pfam: PF00054; laminin_G; 3.	
DR	PRINTS; PR00290; KAZALINBFR.	
DR	PROSITE; PS00010; ASX_HYDROXYD; 1.	
DR	PROSITE; PS00022; EGF_1; 6.	
DR	PROSITE; PS01186; EGF_2; 1.	
DR	PROSITE; PS01248; LAMININ_Type_EGF; 1.	
KM	Glycoprotein; EGF-like domain; Repeat; Alternative splicing; Signal;	
KW	Laminin EGF-like domain.	
FT	SIGNAL	1
FT	CHAIN	39
FT	DOMAIN	54
FT	DOMAIN	130
FT	DOMAIN	202
FT	DOMAIN	276
FT	DOMAIN	350
FT	DOMAIN	419
FT	DOMAIN	484
FT	DOMAIN	551
FT	DOMAIN	675
FT	DOMAIN	729
FT	DOMAIN	781
FT	DOMAIN	851
FT	DOMAIN	856
FT	DOMAIN	1150
FT	DOMAIN	1229
FT	DOMAIN	1446
FT	DOMAIN	1485
FT	DOMAIN	1485
FT	DOMAIN	1714
FT	DISULFID	86
FT	DISULFID	105
FT	DISULFID	94
FT	POTENTIAL.	38
FT	POTENTIAL.	1955
FT	POTENTIAL.	126
FT	POTENTIAL.	201
FT	POTENTIAL.	273
FT	POTENTIAL.	344
FT	POTENTIAL.	418
FT	POTENTIAL.	483
FT	POTENTIAL.	548
FT	POTENTIAL.	633
FT	POTENTIAL.	728
FT	POTENTIAL.	775
FT	POTENTIAL.	851
FT	POTENTIAL.	856
FT	POTENTIAL.	1150
FT	POTENTIAL.	1229
FT	POTENTIAL.	1446
FT	POTENTIAL.	1485
FT	POTENTIAL.	1485
FT	POTENTIAL.	1714
FT	POTENTIAL.	86
FT	POTENTIAL.	105
FT	POTENTIAL.	94
FT	POTENTIAL.	126

FT	DISULFID	160	180	POTENTIAL.
FT	DISULFID	169	201	POTENTIAL.
FT	DISULFID	233	252	POTENTIAL.
FT	DISULFID	241	273	POTENTIAL.
FT	DISULFID	304	323	POTENTIAL.
FT	DISULFID	312	344	POTENTIAL.
FT	DISULFID	378	397	POTENTIAL.
FT	DISULFID	386	418	POTENTIAL.
FT	DISULFID	443	462	POTENTIAL.
FT	DISULFID	451	483	POTENTIAL.
FT	DISULFID	507	527	POTENTIAL.
FT	DISULFID	516	548	POTENTIAL.
FT	DISULFID	592	612	POTENTIAL.
FT	DISULFID	601	633	POTENTIAL.
FT	DISULFID	675	687	BY SIMILARITY.
FT	DISULFID	677	694	BY SIMILARITY.
FT	DISULFID	696	705	BY SIMILARITY.
FT	DISULFID	708	726	BY SIMILARITY.
FT	DISULFID	729	741	BY SIMILARITY.
FT	DISULFID	731	748	BY SIMILARITY.
FT	DISULFID	750	759	BY SIMILARITY.
FT	DISULFID	762	773	BY SIMILARITY.
FT	DISULFID	810	830	POTENTIAL.
FT	DISULFID	819	851	POTENTIAL.
FT	DISULFID	1233	1244	BY SIMILARITY.
FT	DISULFID	1238	1253	BY SIMILARITY.
FT	DISULFID	1255	1264	BY SIMILARITY.
FT	DISULFID	1450	1461	BY SIMILARITY.
FT	DISULFID	1455	1471	BY SIMILARITY.
FT	DISULFID	1473	1482	BY SIMILARITY.
FT	DISULFID	1489	1500	BY SIMILARITY.
FT	DISULFID	1494	1510	BY SIMILARITY.
FT	DISULFID	1512	1521	BY SIMILARITY.
FT	DISULFID	1718	1731	BY SIMILARITY.
FT	DISULFID	1725	1740	BY SIMILARITY.
FT	DISULFID	1742	1751	BY SIMILARITY.
FT	CARBOHYD	390	390	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	659	659	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	764	764	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	814	814	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPPLIC	1451	1551	MISSING (IN AGRIN-RELATED PROTEIN 2).
FT	VARSPPLIC	1783	1793	MISSING (IN AGRIN-RELATED PROTEIN 1 AND AGRIN-RELATED PROTEIN 2).
FT	CONFLICT	1129	1131	RTI -> SIL (IN AAA48586).
SO	SEQUENCE	1955 AA;	211411 MW;	B4DBE27C3423581 CRC64;

	Query Match Similarity	8.1%; Score 78;	DB 1; Length 1955;
	Best Local similarity	22.3%;	Pred. No. 22;
Matches	477;	Conservative	24; Mismatches 62; Indels 78; Gaps 9;
OY	5 DFQPVLLHVALNTPPLSGMR--GIRGADFQCFPOARAVGLSTGFRAF-----	49	
Db	1670 DESKILARAANAISTSEYGANVRISIKGVPLRKQHRIASAIVEITFRAHPCTOKRPNFCONGC	1729	
OY	50 -LSSLDLVLIVRRADKGSVPIVNLKDENVLSPSMDSLFSGSQ-----GQLDPPARRIFS	102	
Db	1730 TCSPLRLSLEYECACORG-----FSGAHCKEKVIIIEKAAGDAEALA	1767	
OY	103 FDGR-----DVLRHPAMPQRKSVMHGSDPSCGRHLMEASYCE-TWRTETTGATGO	148	
Db	1768 FDFGRFYMEYNHAVTKTSPALDYPA-----EPSEEKLAKDSNHFEIISKITEAF----	Q 1813	
OY	149 ASSLLSGRLDECKAASCHNSHYVLCIENTSFQM	179	
Db	1814 GLIMWSGGKLER-----SDYLTAIAIDVGFI	1838	
RESULT	5 SYXJL_RAT	SYXJL_RAT	STANDARD:
ID AC Q62910_062911_-089092;			PRT: 1574 AA.
TITLE	30-MAY-2000 (Rel. 39, Created)		

DT 30-MAY-2000 (Rel. 39, last sequence update)  
 DT 01-OCT-2000 (Rel. 40, last annotation update)  
 DE SYNAPTOJANIN 1 (EC 3.1.3.56) (SYNAPTIC INOSITOL-1,4,5-TRISPHOSPHATE 5-  
 DE PHOSPHATASE 1).  
 GN SYNJ.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sclurognathu; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;  
 RX MEDLINE=96149250; PubMed=8552192;  
 RA McPherson P.S., Garcia E.P., Slepnev V.I., David C., Zhang X.,  
 RA Grab D., Sosain W.S., Bauerfeind R., Nemoto Y., De Camilli P.;  
 RT "A presynaptic inositol-5-phosphatase.";  
 RL Nature 379:353-357(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 3).  
 RC TISSUE=Brain;  
 RX MEDLINE=98374013; PubMed=9710239;  
 RA Moscholakaki R., Finan P.M., Radley E., Parker P.J.;  
 RT "Identification and characterisation of a novel splice variant of  
 RT synaptojanin.";  
 RL FEBS Lett. 432:5-8(1998).  
 RN [3]  
 RP ALTERNATIVE SPLICING.  
 RC TISSUE=Brain;  
 RX MEDLINE=96394655; PubMed=8798761;  
 RA Ramjaun A.R., McPherson P.S.;  
 RT "Tissue-specific alternative splicing generates two synaptojanin  
 RT isoforms with differential membrane binding properties.";  
 RL J. Biol. Chem. 271:24856-24861(1996).  
 CC -1- FUNCTION: INOSITOL-5-PHOSPHATASE WHICH HAS A ROLE IN CLATHRIN-  
 CC MEDIATED ENDOCYTOSIS.  
 CC -1- CATALYTIC ACTIVITY: D-MYO-INOSITOL 1,4,5-TRISPHOSPHATE + H(2)O -  
 CC D-MYO-INOSITOL 1,4-BISPHOSPHATE + PHOSPHATE.  
 CC -1- SUBCELLULAR LOCATION: LOCALIZED MAINLY IN THE SOLUBLE FRACTION (BY  
 CC SIMILARITY).  
 CC -1- ALTERNATIVE PRODUCTS: 3 MAIN ISOFORMS; ISOFORM 1/170 KDA (SHOWN  
 CC HERE), ISOFORM 2/145 KDA AND ISOFORM 3/DELTA-SAC; ARE PRODUCED BY  
 CC ALTERNATIVE SPLICING. A STOP CODON IN POSITION 1309 IS SUPPRESSED  
 CC IN ISOFORM 1 DUE TO ALTERNATIVE SPLICING. A 16 AA INSERT IS  
 CC OPTIONALLY PRESENT IN THE THREE MAIN ISOFORMS.  
 CC -1- TISSUE SPECIFICITY: ISOFORM 1 IS FOUND IN NEONATAL BRAIN, AND IN A  
 CC WIDE VARIETY OF ADULT NON-NEURONAL TISSUES. ISOFORM 2 IS EXPRESSED  
 CC PREDOMINANTLY IN THE NEURONS, BUT IS ALSO FOUND IN ALL OTHER  
 CC TISSUES AT MUCH LOWER LEVELS. ISOFORMS 1 AND 2 ARE DETECTED IN THE  
 CC LUNG AND HEART. ISOFORM 1 IS EXPRESSED AT HIGHER LEVELS THAN  
 CC ISOFORM 2 IN THE TESTIS AND LIVER AND BOTH ISOFORMS ARE NOT  
 CC DETECTED IN THE SKELETAL MUSCLE. ISOFORM 3 WITH THE 16-AMINO-ACID  
 CC INSERT IS ONLY FOUND IN THE BRAIN WHILE ISOFORM 3 WITHOUT THE 16-  
 CC AMINO-ACID INSERT IS FOUND IN THE LUNG.  
 CC -1- DEVELOPMENTAL STAGE: AT EMBRYONIC DAY 12 (E12) ONLY ISOFORM 1 IS  
 CC SEEN WHILE AT E16 AND E18 ISOFORMS 1 AND 2 ARE SEEN. IN THE ADULT  
 CC BRAIN EXPRESSION OF ISOFORM 2 INCREASES DRAMATICALLY AS COMPARED  
 CC WITH ITS EXPRESSION IN EMBRYONIC BRAIN WHERE AS ISOFORM 1  
 CC DECREASES TO UNDETECTABLE LEVELS.  
 CC -1- DOMAIN: BINDS TO EPS15 (A CLATHRIN COAT-ASSOCIATED PROTEIN) VIA A  
 CC C-TERMINAL DOMAIN CONTAINING THREE ASN-PRO-PHE (NPF) REPEATS (BY  
 CC SIMILARITY).  
 CC -1- DOMAIN: THE C-TERMINAL PROLINE-RICH REGION MEDIATES BINDING TO A  
 CC VARIETY OF SH3 DOMAIN-CONTAINING PROTEINS INCLUDING AMPHIPHYSIN,  
 CC SH3P4, SH3P8, SH3P13 AND GRB2.  
 CC -1- DOMAIN: SPLICING OF THE SAC1 DOMAIN DOES NOT ALTER THE CATALYTIC  
 CC ACTIVITY OF SYNAPTOJANIN 1.  
 CC -1- SIMILARITY: IN THE CENTRAL SECTION, BELONGS TO THE INOSITOL-1,4,5-  
 CC TRISPHOSPHATE 5-PHOSPHATASE FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 SAC1 DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U45479; AAB60525.1; -  
 DR EMBL: U45479; AAB60526.1; -  
 DR EMBL: AJ066855; CA07267.1; ALT\_TERM.  
 DR InterPro: IPR000300;  
 DR Pfam: PF00783; IPPC; 1.  
 DR PROSITE: PS50102; RRM; 1.  
 KM Hydrolyase; Alternative splicing; Repeat; Endocytosis; RNA-binding;  
 KM Multigene family.  
 FT DOMAIN 1 499  
 FT 500 899  
 FT 894 971  
 FT DOMAIN 894 1574  
 FT 900 1574  
 FT DOMAIN 1033 1036  
 FT 1105 1110  
 FT DOMAIN 1123 1126  
 FT 1537 1545  
 FT DOMAIN 1401 1423  
 FT 1401 1403  
 FT REPEAT 1410 1412  
 FT REPEAT 1421 1423  
 FT REPEAT 1 400  
 FT VARSPLIC 1140 1155  
 FT VARSPLIC 1309 1574  
 FT CONFLICT 588 588  
 FT CONFLICT 1140 1140  
 FT CONFLICT 1156 1156  
 SQ SEQUENCE 1574 AA; 172880 MW; 9DEEC876BABA3ADF CRC64;  
 Query Match 8.0%; Score 77; DB 1; Length 1574;  
 Best Local Similarity 22.6%; Pred. No. 22;  
 Matches 38; Conservative 21; Mismatches 49; Indels 60; Gaps 8;  
 QY 8 PVLHVALNPLPS--GGMGRIGR-----DPCFOQAR-AVLSGFRFLSSRLQ 55  
 DB 648 PSTRDAVDVKTGMGAGTNGKGAVALRMLEHTTSLCFVCSHFAAQSQ----VERNE 702  
 QY 56 DLVSIVRAD-----RGSVPVINKDEVLSPSMDLFSGSQ--- 91  
 DB 703 DVEFIARKLSFPGKMLFSDHYFMCGDPRYRIDLNEEYKEILROONDLSLAGQLIN 762  
 QY 92 ----GQLQPG-----ARIFS--FDGRDYLRRHPAMPQKSVW 120  
 DB 763 QKNAGQIFRGFLGKGYTFAPTYKDYLFSEDDYDSEKCRFPAMPDRLVM 810  
 RESULT 6  
 ULK1.HUMAN  
 ID ULK1.HUMAN STANDARD; PRT; 1050 AA.  
 AC 075385;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE SERINE/THROMBIN-PROTEIN KINASE ULK1 (EC 2.7.1.-) (UNC-51-LIKE KINASE  
 DE 1).  
 GN ULK1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98360094; PubMed=9693035;  
 RA Kuroyanagi H., Yan J., Seki N., Yamanouchi Y., Suzuki Y.I., Takano T.,  
 RA Muramatsu M.A., Shirasawa T.;  
 RT "Human ULK1, a novel serine/threonine kinase related to UNC-51 kinase  
 RT of Caenorhabditis elegans: cDNA cloning, expression, and chromosomal



```

RT assignment.*
RL Genomics 51:76-85(1998).
CC -1- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED. DETECTED IN THE
CC FOLLOWING ADULT TISSUES: SKELETAL MUSCLE, HEART, PANCREAS, BRAIN,
CC PLACENTA, LIVER, KIDNEY, AND LUNG.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC APG1/UNC-51/DUK1 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF045458; AAC32326.1; -.
DR MIM: 603168; -.
DR InterPro: IPR000719; -.
DR InterPro: IPR001245; -.
DR InterPro: IPR002290; -.
DR Pfam: PF00069; PKinase; 1.
DR PRINTS: PR00109; TYRKINASE.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR Transferrase; Serine/threonine-protein kinase; ATP-binding.
KW DOMAIN 16 278
FT NP_BIND 22 30 ATP (BY SIMILARITY).
FT BINDING 46 46 ATP (BY SIMILARITY).
FT ACT_SITE 138 138 BY SIMILARITY.
FT DOMAIN 297 310 POLY-SER.
SQ SEQUENCE 1050 AA; 112601 MW; 4ED9B94028E3C138 CRC64;

Query Match
Best Local Similarity 7.9%; Score 76.5; DB 1; Length 1050;
Matches 43; Conservative 13; Mismatches 47; Indels 39; Gaps 9;

OY 18 PLSGMRGIRGADFOCFQARAVGLSGTF-RAFLSLRDLDSIVRA-----DGSVYP 70
DB 674 PLGPGLR--FGED-----PKGPGRGSTSTSLTDL--LTKAAQTQAPDPCSTP 718
OY 71 IYNLKDEVLSPSWDSLFSGSGOLOPGARIFSPDGRVLRHPAMPQKSVYH-GSDPSGR 129
DB 719 SIQEKPMETIAPS-----AGFGSLHRCAGARAGTSS-----PSPVVFYTGSPSGST 764
OY 130 LMEISYCETRTETGATGQASS 151
DB 765 PPOG--PRTMFMSAGPTGSASS 784

RESULT 7
YHCC_ECOLI STANDARD: PRT: 309 AA.
AC P43476;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 34.6 KDA PROTEIN IN ARCB-GLTB INTERGENIC REGION.
GN YHCC.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RA MEDLINE-97426617; PubMed-9278503;
RA Biltner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley J.M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;

```

```

RT *The complete genome sequence of Escherichia coli K-12.*
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE OF 1-39 FROM N.A.
RA Erstling B.R., Drenthinger J., Blumenthal R.M., Matthews R.G.;
RL Submitted (JUN-1993) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: TO M.JANNASCHII M10486.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U18997; AAA58013.1; ALT_SEQ.
DR EMBL: AE000400; AAC76243.1; -.
DR EMBL: L20253; -: NOT_ANNOTATED_CDS.
DR Ecogene: EC12809; yhcC.
KW Hypothetical protein.
FT CONFLICT 24 25 KL-> NV (IN REF. 2).
SQ SEQUENCE 309 AA; 34606 MW; 61B3187B77CA1A9 CRC64;

Query Match
Best Local Similarity 7.9%; Score 76; DB 1; Length 309;
Matches 35; Conservative 23; Mismatches 42; Indels 34; Gaps 10;

OY 27 RGADFOCFQD---AR-----AVGLSGTFRAFLSLRDLDSIVRADRGSV--P 70
DB 159 RGHDFPCYQRTTGLARQGLKVCSHLIVGLPGSGQ---ECLDTLRVYETGVGDKLKP 215
OY 71 IYNLKDEVLSPSWDSLFSGSGOLOPGARIFSPDGRVLRHPAMPQKSVYH-GSDPSGR 128
DB 216 LHIYKGSIMAKWEA-----GRLN-GIELEDYTLTAGEMIRHT--PPEVIYHISASAR 266
OY 129 R--LMEISYCET-W 138
DB 267 RPTLLAPLWCENRW 280

RESULT 8
SYJ1_HUMAN STANDARD: PRT: 1575 AA.
AC O43426; O43425;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE SYNAPTOJANIN 1 (EC 3.1.3.56) (SYNAPTIC INOSITOL-1,4,5-TRISPHOSPHATE 5-
DE PHOSPHATASE 1).
GN SYNJ1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum.
RA MEDLINE-98088905; PubMed-9428629;
RA Haffner C., Takei K., Chen H., Ringstad N., Hudson A., Butler M.H.,
RA Salcini A.E., Di Fiore P.P., De Camilli P.;
RA "Synaptojanin 1: localization on coated endocytic intermediates in
RT nerve terminals and interaction of its 170 kDa isoform with Eps15";
RL FEBS Lett. 419:175-180(1997).
CC -1- FUNCTION: INOSITOL 5-PHOSPHATASE WHICH HAS A ROLE IN CLATHRIN-
CC MEDIATED ENDOCYTOSIS.
CC -1- CATALYTIC ACTIVITY: D-MYO-INOSITOL 1,4,5-TRISPHOSPHATE + H(2)O -
CC D-MYO-INOSITOL 1,4-BISPHOSPHATE + PHOSPHATE.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG ISOFORM/SYNAPTOJANIN-
CC 170 (SHOWN HERE) AND A SHORT ISOFORM/SYNAPTOJANIN-145; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: CONCENTRATED AT CLATHRIN-COATED ENDOCYTIC

```

```

CC INTERMEDIATES IN NERVE TERMINALS. THE LONG ISOFORM IS MORE
CC ENRICHED THAN THE SHORT ISOFORM IN DEVELOPING BRAIN AS WELL AS
CC NON-NEURONAL CELLS. THE SHORT ISOFORM IS VERY ABUNDANT IN NERVE
CC TERMINALS.
CC -1- DOMAIN: BINDS TO EPS15 (A CLATHRIN COAT-ASSOCIATED PROTEIN) VIA A
CC C-TERMINAL DOMAIN CONTAINING THREE ASN-PRO-PHE (NPF) REPEATS.
CC -1- DOMAIN: THE C-TERMINAL, PROLINE-RICH REGION MEDIATES BINDING TO A
CC VARIETY OF SH3 DOMAIN-CONTAINING PROTEINS INCLUDING AMPHIPHYLIN,
CC SH3P4 AND GRB2.
CC -1- SIMILARITY: IN THE CENTRAL SECTION, BELONGS TO THE INOSITOL-1,4,5-
CC TRISPHOSPHATE 5-PHOSPHATASE FAMILY.
CC -1- SIMILARITY: CONTAINS 1 SAC1 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC .....
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC .....
CC EMBL: AF009040; AAC51922.1; -
CC EMBL: AF009039; AAC51921.1; -
CC MIM: 604297; -
CC DR InterPro: IPR000300; -
CC DR Pfam: PF00783; IPPE; 1.
CC DR PROSITE: PS50102; RRM; 1.
CC KM Hydrolyase; Alternative splicing; Repeat; Endocytosis; RNA-binding;
CC Multigene family.
CC FT DOMAIN 1 499 SAC1.
CC FT DOMAIN 500 899 CATALYTIC (POTENTIAL).
CC FT DOMAIN 902 971 RNA-BINDING (RRM).
CC FT DOMAIN 900 1575 PRO-RICH.
CC FT DOMAIN 1033 1036 POLY-SER.
CC FT DOMAIN 1108 1113 POLY-PRO.
CC FT DOMAIN 1126 1129 POLY-PRO.
CC FT DOMAIN 1487 1490 POLY-GLU.
CC FT DOMAIN 1540 1546 POLY-PRO.
CC FT DOMAIN 1396 1419 3 X 3 AA REPEATS OF N-P-F.
CC FT REPEAT 1396 1398 1.
CC FT REPEAT 1406 1408 2.
CC FT REPEAT 1417 1419 3.
CC FT REPEAT 1306 1311 VKRNGI -> QEQPSG (IN SHORT ISOFORM).
CC FT VARSPLIC 1312 1575 MISSING (IN SHORT ISOFORM).
CC SO SEQUENCE 1575 AA; 173345 MW; 5064666CC043B9E7 CRC64;

Query Match 7.9%; Score 76; DB 1; Length 1575;
Best Local Similarity 22.6%; Pred. No. 27;
Matches 38; Conservative 21; Mismatches 49; Indels 60; Gaps 8;

QY 8 PVLHLVALNTPLS--GGMRCIRGA-----DFOCFOQAR-AVIGSGFRFRAFLSSRLQ 55
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 648 PFIIDVAVDYKTCMGATGKGAIVAILRMLEHTTSLCFVCSHFAGDSQ-----VKERNE 702
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 56 DLVSIVRRAD-----RCGSPVINKDEVLSPMSDLFSGSO--- 91
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 703 DFEIARKLSPPMGRLMFSHDYVFWCGDFNTRIDLPMNEVEKELRLQGNMDSLLINGDDQLIN 762
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 92 ----GQLQPG-----ARIFS--FDGRDVLRRHPAPOKSW 120
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 763 OKNAGVFRGFLGKVTFAPTYKYDLFSDDYDTSEKCRTPAMTRVLW 810
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 9
CATAL_METBA STANDARD; PRT; 505 AA.
AC 093662;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CATALASE (EC 1.11.1.6).

```

```

GN KAT.
OS Methanosarcina barkeri.
OC Archaea; Euryarchaeota; Methanosarcinales; Methanosarcinaceae;
OC Methanosarcina.
OX NCBI_TaxID=2208;
RN (1)
RP SEQUENCE FROM N.A.
RA STRAIN=FUSARO / DSM 804;
RA Shima S., Netrusov A., Sordel M., Wicke M., Hartmann G.C.,
RA Thauer R.K.;
RT *Purification, characterization and primary structure of a
RT monofunctional catalase from Methanosarcina barkeri.*;
RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: DECOMPOSES HYDROGEN PEROXIDE IN WATER AND OXYGEN; SERVES
CC TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN PEROXIDE.
CC -1- CATALYTIC ACTIVITY: 2 H2O2 = O2 + 2 H2O.
CC -1- COFACTOR: HEME GROUP (BY SIMILARITY)
CC -1- SIMILARITY: BELONGS TO THE CATALASE FAMILY.
CC .....
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC .....
CC EMBL: AJ005939; CAA06774.1; -
CC HSSP: P00432; 7CAT.
CC DR InterPro: IPR002226; -
CC DR Pfam: PF00199; catalase; 1.
CC DR PRINTS: PR00067; CATALASE.
CC DR PROSITE: PS00437; CATALASE_1; 1.
CC DR PROSITE: PS00438; CATALASE_2; 1.
CC KM Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.
CC FT ACT_SITE 58 58 BY SIMILARITY.
CC FT ACT_SITE 131 131 PROXIMAL HEME LIGAND (BY SIMILARITY).
CC FT BINDING 341 341 2A27C4BEC47BE854 CRC64;
CC SO SEQUENCE 505 AA; 57065 MW; 2A27C4BEC47BE854 CRC64;

Query Match 7.7%; Score 75; DB 1; Length 505;
Best Local Similarity 20.7%; Pred. No. 9;
Matches 30; Conservative 30; Mismatches 47; Indels 38; Gaps 7;

QY 35 QGARAVGLSGTFRAFLSSRLQDLYSIYVRADGSPVYNLKDEVLSPMSDLFSGSGQL 94
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 231 EEAEKIGSDPDHA-----TRDLYEAIK---KQDYPSWTLEMQIMPE----- 270
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 95 QPGARIFSPGDRVLRHPAPOKSWHGSDFS--GRLMESYCEWRTETTGATGOASS 151
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 271 --QAEDYRFDIRDTYK--VMP-----HGDEPTMIGKLVLRNPTNFAFEQAASFSPAN 321
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 152 LLSG-----RLLEOKAASCHNSYI 170
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 322 LVPGIGISPDKMLQGRVFSYHDTTHI 346
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 10
HEMO_HUMAN STANDARD; PRT; 587 AA.
ID HEMO_HUMAN
AC P22557;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE 5-AMINOLEVULINIC ACID SYNTHASE, ERYTHROID-SPECIFIC, MITOCHONDRIAL
DE PRECURSOR (EC 2.3.1.37) (DELTA-AMINOLEVULINATE SYNTHASE) (DELTA-ALA
DE SYNTHETASE) (ALAS-E).
GN ALAS2 OR ALASE OR ASB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;

```

Disease mutation. 249  
 KM TRANSIT 1  
 FT CHAIN 750 587  
 FT BINDING 391 391  
 FT VARIANT 388 388  
 FT VARIANT 411 411  
 FT VARIANT 476 476  
 FT SEQUENCE 587 AA; 64693 MM; DIA93AABE21CC794 CRC64;  
 Query Match 7.7%; Score 75; DB 1; Length 587;  
 Best Local Similarity 24.6%; Pred. No. 11;  
 Matches 49; Conservative 23; Mismatches 57; Indels 70; Gaps 10;

OY 49 FLSSLDLQDLYS-IVRRARGSVPIV-----NLKDEVLSPSWDSLFSSGOLQ--- 95  
 :|::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
 Db 72 FMLSLELDQKSKIYKA-----APEVEDYVKAFKTDLPSSLVSILRPFSGPOEDQLSG 127  
 OY 96 -----PGARIFSEDG--RDYLNRH-----PAMP----- 115  
 Db 128 KVTLLIQQNMMPGNYVFSDYDFFRDKIMEKKODHTRYFKEVYNRNADAPFGQHFEASVA 187  
 OY 116 --QKSVHGSDPSGR----RLMESICEWREFTTGATGQASSLSGLR-----LEOKRAS 164  
 :|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
 Db 188 SKDVSVWCNSNDYLGSRHPVOLNQETLRHQAGAGCTRN--ISGTSKFHVELEOELAE 245  
 OY 165 CHNNTYVCICENSPETSFS 183  
 :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
 Db 246 LHKDSALLFFSCFVANDS 264

RESULT 11  
 SYJL\_BOVIN STANDARD; PRT: 1324 AA.  
 ID SYJL\_BOVIN STANDARD;  
 AC O18964;  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE SYNAPTOJANIN 1 (EC 3.1.3.56) (SYNAPTIC INOSITOL-1,4,5-TRISPHTPHATE 5-  
 DE PHOSPHATASE 1) (P150) (FRAGMENT).  
 GN SYNJ1.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 321-339 AND 454-469.  
 RC TISSUE-Brain:  
 RX MEDLINE=97342621; PubMed=9199318;  
 RA Sakisaka T., Itoh T., Miura K., Takenawa T.;  
 RT "Phosphatidylinositol 4,5-bisphosphate phosphatase regulates the  
 rearrangement of actin filaments."  
 RL Mol. Cell. Biol. 17:3841-3849(1997).  
 CC -!- FUNCTION: HYDROLYSES PI2P BOUND TO ACTIN REGULATORY PROTEINS  
 RESULTING IN THE REARRANGEMENT OF ACTIN FILAMENTS DOWNSTREAM OF  
 TYROSINE KINASE AND ASH/GRB2.  
 CC -!- CATALYTIC ACTIVITY: D-MYO-INOSITOL 1,4,5-TRISPHTPHATE + H(2)O =  
 CC D-MYO-INOSITOL 1,4-BISPHTPHATE + PHOSPATE.  
 CC -!- SUBUNIT: BINDS TO AMPHIPHYLIN AND ASH/GRB2.  
 CC -!- SUBCELLULAR LOCATION: PREDOMINANTLY CONCENTRATED IN THE  
 PERINUCLLEAR AREAS.  
 CC -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED WITH HIGHEST LEVELS IN  
 BRAIN.  
 CC -!- DOMAIN: THE C-TERMINAL PROLINE-RICH REGION MEDIATES BINDING TO A  
 CC VARIETY OF SH3 DOMAIN-CONTAINING PROTEINS INCLUDING AMPHIPHYLIN,  
 CC AND ASH/GRB2.

-!- SIMILARITY: IN THE CENTRAL SECTION: BELONGS TO THE INOSTOL-1,4,5-



	Matches	53;	Conservative	40;	Mismatches	77;	Indels	86;	Gaps	14.
QY	4	ODFOFVL-----	HVALNTPLSGGRRGRGADFOCFQOAR-							38
Db	154	ENFERILVESRRMGCVSFAQLTFP--GGV----	LNKEFQARAANAOKLETLTWOPFI							206
QY	39	----AVGLSGTFR-----	FLSSRLQDLY-SIVRADRGSVPIVNLKDE- 77							
Db	207	OGMVAAMCASGCIKAHAHEVLMEMGEKDGIIPPERLEKLVKELVRNRNFSLSLPGISEER								266
QY	78	-----VLSPSDSL-----	FSSGGOLOPGARIFFSDGR---DYLRHPAMPQKSYW 120							
Db	267	KTVFPVGLAIIQGVAFDALAIRLRLSDALRQGV-LYEMEGFRHQDVRSRTASSLANQY								325
QY	121	HGSDPSGRRLMES--YEETWTETTGT-ATGCASSLS-----	GRLECKAASCHNS 168							
Db	326	HIDSQARRVLDITWMQTEOWREQQPKIAPHOLEALLRWAAMLHEVGINIHSGLIHRBSA								385
QY	169	IYVICIENSFMTSFK 184								
Db	386	YI---LQNSDLPGFNQ 398								
	RESULT. 14									
ID	P47K_PSECL	STANDARD;	PRT; 419 AA.							
AC	P31521:									
DT	01-JUL-1993 (Rel. 26, Created)									
DT	01-JUL-1993 (Rel. 26, Last sequence update)									
DT	30-MAY-2000 (Rel. 39, Last annotation update)									
DE	47 KDA PROTEIN (P47K).									
CC	Pseudomonas chlororaphis (Pseudomonas fluorescens biotype D).									
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;									
OX	Pseudomonas									
NCBI_TaxID=333;	[1]									
RN	SEQUENCE FROM N.A.									
RP	STRAIN-B23;									
RC	MEDLINE-91193202; PubMed-2013568;									
RA	Nishiyama M., Horiouchi S., Kobayashi M., Nagasawa T., Yamada H.,									
RA	Beppu T.;									
RT	"Cloning and characterization of genes responsible for metabolism of									
RL	nitrile compounds from Pseudomonas chlororaphis B23.";									
RL	J. Bacteriol. 173:2465-2472(1991).									
CC	-!- FUNCTION: THE PRESENCE OF P47K IS CRITICAL FOR THE EXPRESSION									
CC	OF THE NITRILE HYDRATASE GENES. MAY STABILIZE OR ACTIVATE THE									
CC	NITRILE HYDRATASE PROTEINS.									
CC	-----									
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration									
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -									
CC	the European Bioinformatics Institute. There are no restrictions on its									
CC	use by non-profit institutions as long as its content is in no way									
CC	modified and this statement is not removed. Usage by and for commercial									
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/									
CC	or send an email to license@lsb.sib.ch).									
CC	-----									
DR	EMBL: D90216; BAA1247.1; -									
DR	PIR: D42725; D42725.									
SQ	SEQUENCE 419 AA; 46666 MW; FF5113B00E27FFQC CRC64;									
	Query Match	7.5%; Score 72.5; DB 1; Length 419;								
	Best Local Similarity	22.1%; Pred. No. 13;								
	Matches 40; Conservative 24; Mismatches 60; Indels 57; Gaps 8;									
QY	26	IRGADFQCFQOARAVAGISGFTRAFLLSR-LQDLYSIYRRADRGSVPIYNLKDEVLPSCMD 84								
Db	141	VDSQGQFALLESTDTVARADTFEAHTSTRHLADL-LIEQVEYANVIYLNKKDDLIDEPOYO 198								
QY	85	SLTSSGSOGLQPGARIFS-----FDGRDYLRIHPAMPQK-----SYWHGS 123								
Db	199	AVHAILING-LNPFSARIMPMAHGNAVLTSLDTHLFDPISLAASPQMNRKMKEATDTGPASES 257								

OY 124 DPGS-----RRLMESYCEWRTTGTGATGASLSLGRLLPORA-----ASC 165  
 DB 258 DTGVTSMVYRERAPFHPQNLLEFLQKPMH-----NGRLRSKGYFWLASR 303  
 OY 166 H 166  
 DB 304 H 304

RESULT 15  
 VNUC\_EBOG4  
 ID VNUC\_EBOG4 STANDARD: PRT: 739 AA.  
 AC Q9QCE9;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE NUCLEOPROTEIN (NUCLEOCAPSID PROTEIN).  
 GN NP.  
 OS Ebola virus (strain Gabon-94) (Ebo).  
 OC viruses: ssRNA negative-strand viruses; Mononegavirales; Filoviridae;  
 OC Filovirus.  
 OX NCBL\_TaxID-128947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-99036017; PubMed-9820131;  
 RA Prehaud C.J.C., Hellebrand E., Coudrier D., Volchkov V.E.,  
 RA Volchkova V.A., Feldmann B., Le Guenno B., Bouloy M.,  
 RT Recombinant Ebola virus nucleoprotein and glycoprotein (Gabon 94  
 strain) provide new tools for the detection of human infections.";  
 RL J. Gen. Virol. 79:2565-2572(1998).  
 CC -I- FUNCTION: RESPONSIBLE FOR ENCAPSIDATION OF GENOMIC RNA.  
 CC -I- DOMAIN: THIS PROTEIN CAN BE DIVIDED INTO A HYDROPHOBIC N-TERMINAL  
 CC HALF, AND A HYDROPHILIC AND HIGHLY ACIDIC C-TERMINAL HALF.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; Y09358; CAA70541.1; -  
 KM Nucleoprotein.  
 SO SEQUENCE 739 AA; 83240 MW; 33A96720FBC638E CRC64;

Query Match 7.5%; Score 72.5; DB 1; Length 739;  
 Best Local Similarity 27.0%; Pred. No. 24;  
 Matches 43; Conservative 16; Mismatches 67; Indels 33; Gaps 8;

OY 22 GMRGIRGAD-----FOCFQARAVGLSGTFRAPLSRLD-----LYSIVRADRGSV 69  
 DB 218 GMHVVAGHDANDAVISNSVAQAPSGLL-IVKTVLDHILQKTQGVRLHPLARTA----- 271  
 OY 70 PIVNLKDEV--LSPSWDLFGSGOLOPFCARIFSPGCRDVLRRHPAMPKRS-----VW 120  
 DB 272 ---KVKNEVNSLKAALSL--AKHGEYAFPARLNLGCVNNLEHGLFPQLSAIALGVATA 326  
 OY 121 HGSDPSGRRLMESYCEWRTTGTGATGASLSLGRLLP 159  
 DB 327 HGSTLAGVNVGEQY-QQLREAAATEAKQLOQYAESRELD 364

Search completed: August 28, 2001, 12:52:40  
 Job time: 109 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2001, 12:49:31 ; Search time 12.29 Seconds  
(without alignments)  
308.269 Million cell updates/sec

Title: US-09-589-777a-2

Perfect score: 968  
Sequence: 1 HTHODFQVHLVALNTPLS.....CHNSYVLCIENSFMFSK 184

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: Issued\_Patents\_AA:\*  
2: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	965	99.7	195	1 US-08-159-784-2	Sequence 2, Appl
2	916	94.6	185	3 US-08-985-526-36	Sequence 36, Appl
3	840	86.8	183	4 US-09-206-059-2	Sequence 2, Appl
4	550	56.8	191	1 US-08-159-784-3	Sequence 3, Appl
5	150	15.5	35	3 US-09-046-985-2	Sequence 2, Appl
6	150	15.5	35	4 US-09-474-743-2	Sequence 2, Appl
7	108	11.2	20	2 US-08-740-168A-1	Sequence 1, Appl
8	108	11.2	20	4 US-09-349-429-1	Sequence 1, Appl
9	101	10.4	16	4 US-09-385-442-32	Sequence 32, Appl
10	101	10.4	22	3 US-09-046-985-7	Sequence 7, Appl
11	101	10.4	22	4 US-09-474-743-7	Sequence 7, Appl
12	75	7.7	190	3 US-09-046-985-15	Sequence 15, Appl
13	75	7.7	190	4 US-09-474-743-15	Sequence 15, Appl
14	75	7.7	587	2 US-08-871-266B-18	Sequence 18, Appl
15	75	7.7	587	2 US-09-018-864A-18	Sequence 18, Appl
16	75	7.7	587	2 US-08-871-267B-24	Sequence 24, Appl
17	74.5	7.7	577	2 US-08-756-317-13	Sequence 13, Appl
18	71.5	7.4	1214	2 US-08-231-193A-54	Sequence 54, Appl
19	71.5	7.4	1214	2 US-08-486-273A-54	Sequence 54, Appl
20	71.5	7.4	1214	3 US-08-480-474-54	Sequence 54, Appl
21	71.5	7.4	1214	3 US-08-940-086A-54	Sequence 54, Appl
22	71.5	7.4	1219	2 US-08-231-193A-50	Sequence 50, Appl
23	71.5	7.4	1219	2 US-08-486-273A-50	Sequence 50, Appl
24	71.5	7.4	1219	3 US-08-480-474-50	Sequence 50, Appl
25	71.5	7.4	1219	3 US-08-940-086A-50	Sequence 50, Appl
26	71.5	7.4	1231	2 US-08-231-193A-48	Sequence 48, Appl
27	71.5	7.4	1231	2 US-08-486-273A-48	Sequence 48, Appl

28	71.5	7.4	1231	3 US-08-480-474-48	Sequence 48, Appl
29	71.5	7.4	1231	3 US-08-940-086A-48	Sequence 48, Appl
30	71.5	7.4	1236	2 US-08-231-193A-6	Sequence 6, Appl
31	71.5	7.4	1236	2 US-08-486-273A-6	Sequence 6, Appl
32	71.5	7.4	1236	3 US-08-480-474-6	Sequence 6, Appl
33	71.5	7.4	1236	3 US-08-940-086A-6	Sequence 6, Appl
34	71.5	7.4	1239	2 US-08-231-193A-52	Sequence 52, Appl
35	71.5	7.4	1239	2 US-08-486-273A-52	Sequence 52, Appl
36	71.5	7.4	1239	3 US-08-480-474-52	Sequence 52, Appl
37	71.5	7.4	1239	3 US-08-940-086A-52	Sequence 52, Appl
38	71.5	7.4	1244	2 US-08-231-193A-46	Sequence 46, Appl
39	71.5	7.4	1244	2 US-08-486-273A-46	Sequence 46, Appl
40	71.5	7.4	1244	3 US-08-480-474-46	Sequence 46, Appl
41	71.5	7.4	1244	3 US-08-940-086A-46	Sequence 46, Appl
42	71	7.3	1289	1 US-07-876-280-4	Sequence 4, Appl
43	71	7.3	1289	1 US-07-675-170-4	Sequence 4, Appl
44	71	7.3	1289	1 US-08-063-170-4	Sequence 4, Appl
45	71	7.3	1289	1 US-08-158-232-4	Sequence 4, Appl

#### ALIGNMENTS

RESULT 1  
US-08-159-784-2  
: Sequence 2, Application US/08159784  
: Patent No 5643783  
: GENERAL INFORMATION:  
: APPLICANT: Bjorn R. Olsen  
: TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF  
: NUMBER OF SEQUENCES: 9  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Fish & Richardson  
: STREET: 225 Franklin Street  
: CITY: Boston  
: STATE: Massachusetts  
: COUNTRY: U.S.A.  
: ZIP: 02110-2804  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
: COMPUTER: IBM PS/2 Model 502 or 555X  
: OPERATING SYSTEM: MS-DOS (Version 5.0)  
: SOFTWARE: WordPerfect (Version 5.1)  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/159, 784  
: FILING DATE: December 1, 1993  
: CLASSIFICATION: 530  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER:  
: FILING DATE:  
: ATTORNEY/AGENT INFORMATION:  
: NAME: John F. Freeman  
: REGISTRATION NUMBER: 29, 066  
: REFERENCE/DOCKET NUMBER: 00246/170001  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (617) 542-5070  
: TELEFAX: (617) 542-8906  
: TELEX: 200154  
: INFORMATION FOR SEQ ID NO: 2:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 195  
: TYPE: amino acid  
: STRANDEDNESS: N/A  
: TOPOLOGY: N/A  
: US-08-159-784-2

Query Match 99.7%: Score 965; DB 1; Length 195;  
Best Local Similarity 99.5%: Pred. No. 5.4e-112;  
Matches 183; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
1 HTHODFQVHLVALNTPLSGGMRGIRGADFQCOARAVGISTFRFLSSRLDLYSI 60  
|||||

Db 12 HTHQDFQVYLHLVALNTPLSGMGRGIRGADFCQFOQARAVGLSGTFRAPLSSRLDLYSI 71  
QY 61 VRRADGSPVIVLKDDEVLSFSGSGOLOHCAITFSFDGVDVLRHPAMPQKSV 120  
Db 72 VRRADGSPVIVLKDDEVLSFSGSGOLOHCAITFSFDGVDVLRHPAMPQKSV 131  
QY 121 HGSDFSGRRLMESYCEWTWRTETTGATGOASSLSLGRLLLEOKAASCHNSYIVLCIENSFMT 180  
Db 132 HGSDFSGRRLMESYCEWTWRTETTGATGOASSLSLGRLLLEOKAASCHNSYIVLCIENSFMT 191  
QY 181 SFSK 184  
Db 192 SFSK 195

RESULT 2  
US-08-985-526-36  
Sequence 36 Application US/08985526  
Patent No. 6080728

GENERAL INFORMATION:  
APPLICANT: Mixson, James A  
TITLE OF INVENTION: CARRIER-DNA COMPLEXES CONTAINING DNA  
TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE  
TITLE OF INVENTION: THERAPY  
NUMBER OF INVENTIONS: 43  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Connolly, Bove, Lodge, & Hultz  
STREET: 1220 Market Street, P.O. Box 2207  
CITY: Wilmington  
STATE: Delaware  
COUNTRY: U.S.A.  
ZIP: 19899

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/985,526  
FILING DATE:

CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/608,845  
FILING DATE: 16-JUL-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: McMorrow Jr., Robert G  
TELEPHONE: (302) 658-9141  
TELEFAX: (302) 658-9141

INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 185 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-985-526-36

Query Match 94.6%; Score 916; DB 3; Length 185;  
Best Local Similarity 95.1%; Pred. No. 6.1e-106;  
Matches 176; Conservative 5; Mismatches 2; Indels 2; Gaps 2;

QY 1 HTHQDFQVYLHLVALNTPLSGMGRGIRGADFCQFOQARAVGLSGTFRAPLSSRLDLYSI 60  
Db 2 HTHQDFQVYLHLVALNTPLSGMGRGIRGADFCQFOQARAVGLSGTFRAPLSSRLDLYSI 60  
QY 61 VRRADGSPVIVLKDDEVLSFSGSGOLOHCAITFSFDGVDVLRHPAMPQKSV 119  
Db 61 VRRADGSPVIVLKDDEVLSFSGSGOLOHCAITFSFDGVDVLRHPAMPQKSV 120  
QY 120 HGSDFSGRRLMESYCEWTWRTETTGATGOASSLSLGRLLLEOKAASCHNSYIVLCIENSFMT 179  
Db 121 HGSDFSGRRLMESYCEWTWRTETTGATGOASSLSLGRLLLEOKAASCHNSYIVLCIENSFMT 180

QY 180 TSFSK 184  
Db 181 TSFSK 185

RESULT 3  
US-09-206-059-2  
Sequence 2, Application US/09206059  
Patent No. 6201104

GENERAL INFORMATION:  
APPLICANT: Macdonald, Nicholas  
APPLICANT: Sim, Kim Lee  
TITLE OF INVENTION: Angiogenesis-Inhibiting Protein Binding Peptides and  
TITLE OF INVENTION: Proteins and Methods of Use  
FILE REFERENCE: 05213-0370  
CURRENT APPLICATION NUMBER: US/09/206,059  
CURRENT FILING DATE: 1998-12-04  
NUMBER OF SEQ ID NOS: 80  
SOFTWARE: Patent Ver. 2.0  
SEQ ID NO 2  
LENGTH: 183  
TYPE: PRP  
ORGANISM: Homo sapiens  
US-09-206-059-2

Query Match 86.8%; Score 840; DB 4; Length 183;  
Best Local Similarity 85.6%; Pred. No. 1.7e-96;  
Matches 155; Conservative 15; Mismatches 11; Indels 0; Gaps 0;

QY 1 HTHQDFQVYLHLVALNTPLSGMGRGIRGADFCQFOQARAVGLSGTFRAPLSSRLDLYSI 60  
Db 1 HTHQDFQVYLHLVALNTPLSGMGRGIRGADFCQFOQARAVGLSGTFRAPLSSRLDLYSI 60  
QY 61 VRRADGSPVIVLKDDEVLSFSGSGOLOHCAITFSFDGVDVLRHPAMPQKSV 120  
Db 61 VRRADGSPVIVLKDDEVLSFSGSGOLOHCAITFSFDGVDVLRHPAMPQKSV 120  
QY 121 HGSDFSGRRLMESYCEWTWRTETTGATGOASSLSLGRLLLEOKAASCHNSYIVLCIENSFMT 180  
Db 121 HGSDFSGRRLMESYCEWTWRTETTGATGOASSLSLGRLLLEOKAASCHNSYIVLCIENSFMT 180

QY 181 S 181  
Db 181 A 181

RESULT 4  
US-08-159-784-3  
Sequence 3, Application US/08159784  
Patent No. 5643783

GENERAL INFORMATION:  
APPLICANT: Bjorn R. Olsen  
TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts

COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
COMPUTER: IBM PS/2 Model 502 or 555X  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/159,784  
FILING DATE: December 1, 1993  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:



TOPOLOGY: linear

APPLICANT: O'Reilly, Mi

APPLICANT: Folkman, M. Judah  
TITLE OF INVENTION: Therapeutic Antiangiogenic Compositions  
TITLE OF INVENTION: and Methods  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Jones & Askeew, LLP  
STREET: 191 Peachtree, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/740,168A  
FILING DATE: 22-OCT-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0223  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: Internal  
ORIGINAL SOURCE:  
ORGANISM: Murine  
TISSUE TYPE: Collagen  
US-08-740-168A-1

Query Match 11.2%; Score 108; DB 2; Length 20;  
Best Local Similarity 100.0%; Pred. No. 5.5e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQVPLHLVALNTPLS 20  
DB 1 HTHQDFQVPLHLVALNTPLS 20

RESULT 8  
US-09-349-429-1  
Sequence 1, Application US/09349429  
Patent No. 6174861  
GENERAL INFORMATION:  
APPLICANT: O'Reilly, Michael  
APPLICANT: Folkman, M. Judah  
TITLE OF INVENTION: Therapeutic Antiangiogenic Compositions  
TITLE OF INVENTION: and Methods  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Jones & Askeew, LLP  
STREET: 191 Peachtree, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/349,429  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/740,168  
FILING DATE: 22-OCT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0223  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: Internal  
ORIGINAL SOURCE:  
ORGANISM: Murine  
TISSUE TYPE: Collagen  
US-09-349-429-1

Query Match 11.2%; Score 108; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 5.5e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQVPLHLVALNTPLS 20  
DB 1 HTHQDFQVPLHLVALNTPLS 20

RESULT 9  
US-09-385-442-32  
Sequence 32, Application US/09385442  
Patent No. 6200954  
GENERAL INFORMATION:  
APPLICANT: Ge, Ruowen  
APPLICANT: Kini, R. Manjunatha  
TITLE OF INVENTION: Small Peptides Having Potent Anti-Angiogenic Activity  
FILE REFERENCE: 1781-170P  
CURRENT APPLICATION NUMBER: US/09/385,442  
CURRENT FILING DATE: 1999-08-30  
EARLIER APPLICATION NUMBER: 60/099,313  
EARLIER FILING DATE: 1999-09-04  
NUMBER OF SEQ ID NOS: 50  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 32  
LENGTH: 16  
TYPE: PRT  
ORGANISM: mammalian  
FEATURE:  
OTHER INFORMATION: Endo-4  
US-09-385-442-32

Query Match 10.4%; Score 101; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 2.9e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 HPAMPKSVWHGSDPS 126  
DB 1 HPAMPKSVWHGSDPS 16

RESULT 10  
US-09-046-985-7  
Sequence 7, Application US/09046985  
Patent No. 6121236  
GENERAL INFORMATION:  
APPLICANT: Ben-Sasson, Shmuel A.  
TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Millitia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/046,985  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brook, David E.  
REGISTRATION NUMBER: 22,592  
REFERENCE/DOCKET NUMBER: CMCC-614  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781) 861-9540  
TELEFAX: (781) 861-6240  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1  
OTHER INFORMATION: /label= modified aa  
US-09-046-985-7

Query Match 10.4%; Score 101; DB 3; Length 22;  
Best Local Similarity 100.0%; Pred. No. 4.7e-06;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 TFRFLSSRLDLYSIVRRAD 65  
|||||  
DB 1 TFRFLSSRLDLYSIVRRAD 21

RESULT 11  
US-09-474-743-7  
Sequence 7, Application US/09474743  
Patent No. 6235716  
GENERAL INFORMATION:  
APPLICANT: Ben-Sasson, Shmuel A.  
TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Millitia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/474,743  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/046,985  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brook, David E.  
REGISTRATION NUMBER: 22,592  
REFERENCE/DOCKET NUMBER: CMCC-614  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781) 861-9540  
TELEFAX: (781) 861-6240  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1  
OTHER INFORMATION: /label= modified aa  
US-09-474-743-7

Query Match 10.4%; Score 101; DB 4; Length 22;  
Best Local Similarity 100.0%; Pred. No. 4.7e-06;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 TFRFLSSRLDLYSIVRRAD 65  
|||||  
DB 1 TFRFLSSRLDLYSIVRRAD 21

RESULT 12  
US-09-046-985-15  
Sequence 15, Application US/09046985  
Patent No. 6121236  
GENERAL INFORMATION:  
APPLICANT: Ben-Sasson, Shmuel A.  
TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Millitia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/046,985  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brook, David E.  
REGISTRATION NUMBER: 22,592  
REFERENCE/DOCKET NUMBER: CMCC-614  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781) 861-6240

TELEFAX: (781) 861-9540  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 190 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-046-985-15

Query Match 7.7%; Score 75; DB 3; Length 190;  
Best Local Similarity 44.7%; Pred. No. 0.21;  
Matches 21; Conservative 2; Mismatches 12; Indels 12; Gaps 1;

OY 21 GGMGIRGADPFCFOQARAVGLSGTFRALFSSRLQDLYSTRADRG 67  
DB 34 GSGSGGSKDFT-----AYRWRLSHRPKDLXSTRADRG 68

RESULT 13  
US-09-474-743-15  
Sequence 15, Application US/09474743  
Patent No. 6235716  
GENERAL INFORMATION:  
APPLICANT: Ben-Sasson, Shmuel A.  
TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/474,743  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/046,985  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brook, David E.  
REGISTRATION NUMBER: 22,592  
REFERENCE/DOCKET NUMBER: CMCC-614  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781) 861-6240  
TELEFAX: (781) 861-9540  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 190 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-474-743-15

Query Match 7.7%; Score 75; DB 4; Length 190;  
Best Local Similarity 44.7%; Pred. No. 0.21;  
Matches 21; Conservative 2; Mismatches 12; Indels 12; Gaps 1;

OY 21 GGMGIRGADPFCFOQARAVGLSGTFRALFSSRLQDLYSTRADRG 67  
DB 34 GSGSGGSKDFT-----AYRWRLSHRPKDLXSTRADRG 68

RESULT 14  
US-08-871-266B-18  
Sequence 18, Application US/08871266B  
Patent No. 5871991  
GENERAL INFORMATION:  
APPLICANT: Eliod, Susan L.  
APPLICANT: Cherry, Joel R.  
TITLE OF INVENTION: Aspergillus Oryzae 5-Aminolevulinic  
Acid Synthases And Nucleic Acids Encoding Same  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 58719910 No. 5871991disk Of No. 5871991th America, Inc.  
STREET: 405 Lexington Avenue - 64th Fl.  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10174  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/871,266B  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambitis, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-878-9652  
TELEFAX: 212-878-9655  
TELEX:  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 587 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: NO. 5871991e  
US-08-871-266B-18

Query Match 7.7%; Score 75; DB 2; Length 587;  
Best Local Similarity 24.6%; Pred. No. 1.1;  
Matches 49; Conservative 23; Mismatches 57; Indels 70; Gaps 10;

OY 49 FLSSRLQDLYS-IVRRADRGSPVIV-----NKDEVLSPSWDSLFSGSGGLQ--- 95  
DB 72 FMSLELDGSKSIYOKA-----APEVEDYKAFKTDLPSSLVSLSLRPFSGPQEOEOLSG 127  
OY 96 -----PGARIRSFDC--RDYLRH-----FAMP----- 115  
DB 128 KYVHLIÖNNMNGVVSVDYDFRDKIMEKKQDHTYRVKTVNEMADAYPPAQHFFEAASVA 187  
OY 116 --QKSVWHSQDPGSR-----BLMESYCEFTWRTETTGATGQASLSGLR-----LEOKAAS 164  
DB 188 SKDVSYWCSDNYLQMSNRHPOYLQATQETLDRHAGAGGTNN--ISGTSKHVLEDEQLAE 245  
OY 165 CHNSYIVLCIENSEFMTSFS 183  
DB 246 LHKDSALLFSSCFVANDS 264

RESULT 15  
US-09-018-864A-18  
Sequence 18, Application US/09018864A  
Patent No. 5958747  
GENERAL INFORMATION:  
APPLICANT: Eliod, Susan L.  
APPLICANT: Cherry, Joel R.





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2001, 12:50:31 ; Search time 22.57 Seconds  
(without alignments)  
1078.605 Million cell updates/sec

Title: US-09-589-777A-2

Perfect score: 968  
Sequence: 1 HTHQDFQVPLHLVALNTPLS.....CHNSYIVLCIENSFTSFSK 184

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP\_ARCHAEA:\*  
2: SP\_BACTERIA:\*  
3: SP\_FUNGI:\*  
4: SP\_HUMAN:\*  
5: SP\_INVERTEBRATE:\*  
6: SP\_MAMMAL:\*  
7: SP\_MNC:\*  
8: SP\_ORGANELLE:\*  
9: SP\_PHAGE:\*  
10: SP\_PLANT:\*  
11: SP\_PROTIST:\*  
12: SP\_UNCLASSIFIED:\*  
13: SP\_VERTEBRATE:\*  
14: SP\_VIRUS:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	968	100.0	184	09JRK63	09JRK63 mus musculu
2	968	100.0	1140	061434	061434 mus musculu
3	968	100.0	1774	062001	062001 mus musculu
4	927	95.8	1226	09Q02D2	09Q02D2 rattus norv
5	858	88.6	1771	09QWU5	09QWU5 rattus norv
6	754	77.9	386	093419	093419 gallus gall
7	574	59.3	1367	035206	035206 mus musculu
8	573	59.2	1367	09EQD9	09EQD9 mus musculu
9	564	58.3	1388	09Y4W4	09Y4W4 homo sapien
10	425.5	44.0	581	09VSD9	09VSD9 homo sapien
11	376	38.8	650	017866	017866 caenorhabdi
12	376	38.8	778	09Q9K6	09Q9K6 caenorhabdi
13	376	38.8	1117	09Q9K7	09Q9K7 caenorhabdi
14	85	8.8	995	09Y3M8	09Y3M8 homo sapien
15	83.5	8.6	651	09VFA9	09VFA9 drosophila
16	82	8.5	483	09USL0	09USL0 subteriles d
17	81	8.4	539	09L4Y1	09L4Y1 streptomyce
18	79	8.2	427	062700	062700 rattus norv
19	79	8.2	1112	09YUR6	09YUR6 turkey aden

20	78.5	8.1	416	2	09L609	09L609 salmonella
21	78.5	8.1	9477	2	09L4X3	09L4X3 streptomyce
22	78	8.1	396	5	09YDN3	09YDN3 drosophila
23	77.5	8.1	1055	10	039725	039725 euglena gra
24	77.5	8.0	474	2	050185	050185 mycobacteri
25	77.5	8.0	6420	2	P95814	P95814 streptomyce
26	77.5	7.7	208	2	09X9V1	09X9V1 streptomyce
27	76.5	7.9	235	7	09NNW8	09NNW8 oncorhynch
28	76.5	7.9	784	4	09Q028	09Q028 homo sapien
29	76.5	7.9	1054	2	09Z1X4	09Z1X4 pseudalter
30	76	7.9	1193	4	094984	094984 homo sapien
31	75.5	7.8	1136	10	09LW65	09LW65 oryza sativ
32	75.5	7.8	4924	3	09P7T1	09P7T1 schizosacch
33	75	7.7	587	4	013735	013735 homo sapien
34	74.5	7.7	167	7	046771	046771 bos taurus
35	74.5	7.7	436	5	09YAK0	09YAK0 drosophila
36	74.5	7.7	3670	2	09Z4X5	09Z4X5 streptomyce
37	74.5	7.7	7576	2	09ZGA4	09ZGA4 streptomyce
38	74	7.6	244	2	053235	053235 rhodobacter
39	74	7.6	497	2	P76981	P76981 escherichia
40	74	7.6	517	14	083378	083378 rat leukem
41	74	7.6	571	2	09L0Q4	09L0Q4 streptomyce
42	74	7.6	1715	6	09GLM4	09GLM4 bos taurus
43	74	7.6	4180	2	09I5N6	09I5N6 pseudomonas
44	74	7.6	6396	2	09KID7	09KID7 streptomyce
45	73.5	7.6	305	10	09M3R8	09M3R8 arabidopsis

## ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	184 AA.
09JRK63	09JRK63	09JRK63		
AC	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)			
DE	ENDOSTATIN (FRAGMENT).			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.			
OX	NCBI_Taxid=10090;			
RM	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRATIN-CHINESE KUNMING;			
RA	Jia S., Zhu F., Xing G., Yu Y., Duan C., Xiu R.-J., He F.;			
RT	*Anticancer treatment of targeted fusion protein delivery to tumor			
RT	neovascularization.*			
RL	Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF257775; AAF69009.1; -			
FT	NON-TER			
FT	NON-TER			
SQ	SEQUENCE	184 AA;	20376 MW;	AC06P9D8D103412A CRC64;
Query Match	Best Local Similarity	100.0%;	Score 968;	DB 11;
Matches 184;	Conservative	0;	Mismatches	0;
Indels	0;	Gaps	0;	
QY	1	HTHODFOVPLHLVALNTPLSGMRCIRGADFCFOARAVGLSGFRFLSLRLDLYSI	60	
DB	1	HTHODFOVPLHLVALNTPLSGMRCIRGADFCFOARAVGLSGFRFLSLRLDLYSI	60	
QY	61	VRADRGSPVIVNLKDEVLSPSMDSLFSGSGQQLPGARIFSFDGRDYLRRHPAMPQKSVW	120	
DB	61	VRADRGSPVIVNLKDEVLSPSMDSLFSGSGQQLPGARIFSFDGRDYLRRHPAMPQKSVW	120	
QY	121	HSDPSGRRLMSYCEETRTTGTGATGQASSLSRLLEQRAASHNSYIVLCIENSFT	180	
DB	121	HSDPSGRRLMSYCEETRTTGTGATGQASSLSRLLEQRAASHNSYIVLCIENSFT	180	
QY	181	SFSK 184		

Db 181 SFSK 184

```

RESULT 2
ID 061434 PRELIMINARY; PRT: 1140 AA.
AC 061434;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE COLLAGEN (FRAGMENT).
GN COL15A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-94059075; PubMed-8240330;
RA Abe N., Muragaki Y., Yoshida H., Inoue H., Nimomiyu Y.;
RT Identification of a novel collagen chain represented by extensive
RT interruptions in the triple-helical region.
RL Cell. Mol. Biol. Res. 196:576-582(1993).
DR EMBL: D17546; BAA04483.1; -.
DR HSSP: P39061; IKOE.
DR MGD: MGI:88449; Col15a1.
DR InterPro: IPR000087; -.
DR Pfam: PF01391; Collagen; 6.
FT NON_TER 1
SQ SEQUENCE 1140 AA; 115156 MW; 8B0C7E6862B3BDFE CRC64;

```

Query Match 100.0%; Score 968; DB 11; Length 1140;  
 Best Local Similarity 100.0%; Pred. No. 6.9e-89;  
 Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 HTHQDFQPVHLVLAALNTPLSGKRGIRGADFOCFQOARAVGLSGTFRALSSRLDLYSI 60
DB 957 HTHQDFQPVHLVLAALNTPLSGKRGIRGADFOCFQOARAVGLSGTFRALSSRLDLYSI 1016
QY 61 VRRADGSPVIVLKDENVLSPSWDSLFSGSGQLQPGARIFSGDGVLRHPAMPQKSWA 120
DB 1017 VRRADGSPVIVLKDENVLSPSWDSLFSGSGQLQPGARIFSGDGVLRHPAMPQKSWA 1076
QY 121 HGSDDPSGRRLMESYCEWTRETTGATGQASSLSGRLLLEOKAASCHNSYIVLCIENSFWT 180
DB 1077 HGSDDPSGRRLMESYCEWTRETTGATGQASSLSGRLLLEOKAASCHNSYIVLCIENSFWT 1136
QY 181 SFSK 184
DB 1137 SFSK 1140

```

```

RT Primary structure of the alpha 1 chain of mouse type XVIII collagen,
RT partial structure of the corresponding gene, and comparison of the
RT alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen
RT chain.
RL J. Biol. Chem. 269:13929-13935(1994).
RN [2]
RP SEQUENCE OF 1-562 FROM N.A.
RX MEDLINE-94240112; PubMed-8183894;
RA Rehn M., Philajantemi T.;
RT "Alpha 1(XVIII)", a collagen chain with frequent interruptions in the
RT collagenous sequence, a distinct tissue distribution, and homology
RT with type XV collagen.
RL Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).
RN [3]
RP SEQUENCE OF 1-562 FROM N.A.
RX MEDLINE-95181468; PubMed-7876242;
RA Rehn M., Philajantemi T.;
RT Identification of three N-terminal ends of type XVIII collagen chains
RT and tissue-specific differences in the expression of the corresponding
RT transcripts. The longest form contains a novel motif homologous to rat
RT and Drosophila frizzled proteins.
RL J. Biol. Chem. 270:4705-4711(1995).
DR EMBL: U03715; AAC52903.1; -.
DR EMBL: U03716; AAC52903.1; JOINED.
DR EMBL: U03718; AAC52903.1; JOINED.
DR EMBL: U34607; AAC52903.1; JOINED.
DR EMBL: U34608; AAC52903.1; JOINED.
DR EMBL: U34609; AAC52903.1; JOINED.
DR EMBL: U34610; AAC52903.1; JOINED.
DR EMBL: U34611; AAC52903.1; JOINED.
DR EMBL: U34612; AAC52903.1; JOINED.
DR EMBL: U34613; AAC52903.1; JOINED.
DR EMBL: U11637; AAC52179.1; -.
DR HSSP: P39061; IKOE.
DR MGD: MGI:88451; Col18a1.
DR InterPro: IPR000024; -.
DR InterPro: IPR000087; -.
DR InterPro: IPR001791; -.
DR InterPro: IPR003129; -.
DR Pfam: PF01392; Collagen; 6.
DR Pfam: PF02210; TSPN; 1.
DR SMART: SM00282; LamG; 1.
KW Signal.
SQ SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88EF232 CRC64;

```

Query Match 100.0%; Score 968; DB 11; Length 1774;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-88;  
 Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 HTHQDFQPVHLVLAALNTPLSGKRGIRGADFOCFQOARAVGLSGTFRALSSRLDLYSI 60
DB 1591 HTHQDFQPVHLVLAALNTPLSGKRGIRGADFOCFQOARAVGLSGTFRALSSRLDLYSI 1650
QY 61 VRRADGSPVIVLKDENVLSPSWDSLFSGSGQLQPGARIFSGDGVLRHPAMPQKSWA 120
DB 1651 VRRADGSPVIVLKDENVLSPSWDSLFSGSGQLQPGARIFSGDGVLRHPAMPQKSWA 1710
QY 121 HGSDDPSGRRLMESYCEWTRETTGATGQASSLSGRLLLEOKAASCHNSYIVLCIENSFWT 180
DB 1711 HGSDDPSGRRLMESYCEWTRETTGATGQASSLSGRLLLEOKAASCHNSYIVLCIENSFWT 1770
QY 181 SFSK 184
DB 1771 SFSK 1774

```

```

RESULT 4
ID 090ZD2 PRELIMINARY; PRT: 226 AA.
AC 090ZD2;
DT 01-MAY-2000 (TREMblrel. 13, Created)

```



DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, last annotation update)  
 DE COLLAGEN XVIII (FRAGMENT).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SPRAGUE-DAWLEY;  
 RA Chen L., Perletti G., Folkman J.;  
 RT "Antitumor activity of rat endostatin.";  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF189709; AAF00975.1; -  
 DR HSSP; P39061; IKOE.  
 KW Collagen.  
 FT NON\_TER  
 SQ SEQUENCE 226 AA; 25350 MW; 38883C0486C0E949 CRC64;

Query Match 95.8%; Score 927; DB 11; Length 226;  
 Best Local Similarity 95.7%; Pred. No. 1.2e-85;  
 Matches 176; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY 1 HTHODFQPVHLVALNTPLSGMRGIRGADFOCFQOARAVGLSGTFRAPLSSRLDLYSI 60  
 DB 43 HTHODFHPVHLVALNTPLSGMRGIRGADFOCFQOARAVGLSGTFRAPLSSRLDLYSI 102  
 OY 61 VRRADGSPVIVNLKDEVLSFSGSGQLPGARIFSPDGRVLRHPAMPQKSVW 120  
 DB 103 VRRADGSPVIVNLKDEVLSFSGSGQLPGARIFSPDGRVLRHPAMPQKSVW 162  
 OY 121 HGSDDPSGRRLMESYCEWTETTGATGQASSLSGRLLBOKAASCHNSYIVLCIENSPFW 180  
 DB 163 HGSDDPSGRRLMESYCEWTETTGATGQASSLSGRLLBOKAASCHNSYIVLCIENSPFW 222  
 OY 181 SFSK 184  
 DB 223 SFSK 226

RESULT 5  
 ID 09W0W5 PRELIMINARY; PRT; 171 AA.  
 AC 09W0W5;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, last annotation update)  
 DE COLLAGEN TYPE XVIII, ALPHA (I) CHAIN (FRAGMENT).  
 GN COL18A1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LIVER;  
 RA Jia J.D., Bauer M., Eberspacher U., Donner P., Schuppan D.;  
 RT "Temporal expression of collagen XVIII/endostatin.";  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ236873; CAB44263.1; -  
 DR HSSP; P39061; IKOE.  
 KW Collagen.  
 FT NON\_TER  
 SQ SEQUENCE 171 AA; 18933 MW; 81BE2EE3FC2C8E72 CRC64;

Query Match 88.6%; Score 858; DB 11; Length 171;  
 Best Local Similarity 95.9%; Pred. No. 7.9e-79;  
 Matches 164; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 9 VLHLVALNTPLSGMRGIRGADFOCFQOARAVGLSGTFRAPLSSRLDLYSIYRRADRS 68  
 DB 1 VLHLVALNTPLSGMRGIRGADFOCFQOARAVGLSGTFRAPLSSRLDLYSIYRRADRS 60  
 OY 69 VPIVNLKDEVLSFSGSGQLPGARIFSPDGRVLRHPAMPQKSVWHSDDPSGR 128  
 DB 61 VPIVNLKDEVLSFSGSGQLPGARIFSPDGRVLRHPAMPQKSVWHSDDPSGR 120  
 OY 129 RLMEYCEWTETTGATGQASSLSGRLLBOKAASCHNSYIVLCIENSPFW 179  
 DB 121 RLMEYCEWTETTGATGQASSLSGRLLBOKAASCHNSYIVLCIENSPFW 171

RESULT 6  
 ID 093419 PRELIMINARY; PRT; 386 AA.  
 AC 093419;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, last annotation update)  
 DE COLLAGEN XVIII (FRAGMENT).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Haller W., Dong S., Schurer B., Cole G.;  
 RT "Collagen XVIII is a basement membrane heparan sulfate proteoglycan.";  
 RL J. Biol. Chem. 0:0-0(1998).  
 DR EMBL; AF083440; AAC33294.1; -  
 DR HSSP; P39061; IKOE.  
 FT InterPro; IPR000087; -  
 SQ SEQUENCE 386 AA; 41775 MW; 34D40FA9EBA3B0E CRC64;

Query Match 77.9%; Score 754; DB 13; Length 386;  
 Best Local Similarity 75.5%; Pred. No. 6.6e-68;  
 Matches 139; Conservative 21; Mismatches 24; Indels 0; Gaps 0;

OY 1 HTHODFQPVHLVALNTPLSGMRGIRGADFOCFQOARAVGLSGTFRAPLSSRLDLYSI 60  
 DB 203 HTHODFQPVHLVALNTPLSGMRGIRGADFOCFQOARAVGLSGTFRAPLSSRLDLYSI 262  
 OY 61 VRRADGSPVIVNLKDEVLSFSGSGQLPGARIFSPDGRVLRHPAMPQKSVW 120  
 DB 263 VRRADGSPVIVNLKDEVLSFSGSGQLPGARIFSPDGRVLRHPAMPQKSVW 322  
 OY 121 HGSDDPSGRRLMESYCEWTETTGATGQASSLSGRLLBOKAASCHNSYIVLCIENSPFW 180  
 DB 323 HGSDDPSGRRLMESYCEWTETTGATGQASSLSGRLLBOKAASCHNSYIVLCIENSPFW 382  
 OY 181 SFSK 184  
 DB 383 AAKK 386

RESULT 7  
 ID 035206 PRELIMINARY; PRT; 1367 AA.  
 AC 035206;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)  
 DE PROCOLLAGEN, TYPE XV (TYPE XV COLLAGEN).

```

CN COL15A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97480713; PubMed=9339358;
RA Hagg P.M., Horelli-Kultunen N., Eklund L., Palotie A.,
RA Pihlajaniemi T.;
RT *Cloning of mouse type XV collagen sequences and mapping of the
RT corresponding gene to 4B1-3. Comparison of mouse and human alpha 1
RT (XV) collagen sequences indicates divergence in the number of small
RT collagenous domains".
RL Genomics 43:31-41(1997).
DR EMBL; AF011450; AAC53387.1; -.
DR HSSP; P39061; 1KOE.
DR MGD; MGI:88449; COL15a1.
DR InterPro; IPR000087; -.
DR InterPro; IPR001791; -.
DR InterPro; IPR003129; -.
DR Pfam; PF01391; Collagen; 4.
DR Pfam; PF02210; TSPN; 1.
DR SMART; SM00282; LamG; 1.
SQ SEQUENCE 1367 AA; 140524 MW; A483A1254AF3AEEC CRC64;

Query Match          59.3%; Score 574; DB 11; Length 1367;
Best Local Similarity 62.4%; Pred. No. 4.4e-49;
Matches 111; Conservative 22; Mismatches 41; Indels 4; Gaps 1;

QY 7 0PVLHLVALNTPILSGMRGIRGADFCQFOARAVGLSTGFRAFLSSRLDLYSTVRRADR 66
DB 1194 RPYLHLVALNTPVAGDIR---ADFCFOARAGLSTFRFLSSRLDLYSTVRRADR 1249
QY 67 GSVPIYNLKDENVLSPWSDLSFGSGOQLQPGARIFSPDGRDVLRRHPAPQKSVWHGSDPS 126
DB 1250 FGLPIYNLKGQVLFNWNDSJFSGDGFNTHTPIYSPDGRDVTDPSPQKVVWHGSNPH 1309
QY 127 GRRLMESYCEWTRETGTGATGQASSLSGRLLDEKAAKSNSTIVLCIENSFMTSPSK 184
DB 1310 GVRLVKDYCEAWRTTDMAYVGFASPLSTGKILDKKAYSCANRLIVLCIENSFMTDRK 1367

RESULT 8
Q9EOD9 PRELIMINARY; PRT; 1367 AA.
AC Q9EOD9;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE TYPE XV COLLAGEN.
GN COL15A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97480713; PubMed=9339358;
RA Hagg P.M., Horelli-Kultunen N., Eklund L., Palotie A.,
RA Pihlajaniemi T.;
RT *Cloning of mouse type XV collagen sequences and mapping of the
RT corresponding gene to 4B1-3. Comparison of mouse and human alpha 1
RT (XV) collagen sequences indicates divergence in the number of small
RT collagenous domains".
RL Genomics 45:31-41(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=129/SV;
RA PubMed=11068203;
RA Eklund L., Muona A., Lietard J., Pihlajaniemi T.;

```

```

RT *Structure of the mouse type XV collagen gene, COL15a1, comparison
RT with the human COL15A1 gene and functional analysis of the promoters
RT of both genes.";
RL Matrix Biol. 19:489-500(2000).
DR EMBL; AF261131; AAG27545.1; -.
DR EMBL; AF261109; AAG27545.1; JOINED.
DR EMBL; AF261110; AAG27545.1; JOINED.
DR EMBL; AF261111; AAG27545.1; JOINED.
DR EMBL; AF261112; AAG27545.1; JOINED.
DR EMBL; AF261113; AAG27545.1; JOINED.
DR EMBL; AF261114; AAG27545.1; JOINED.
DR EMBL; AF261115; AAG27545.1; JOINED.
DR EMBL; AF261116; AAG27545.1; JOINED.
DR EMBL; AF261117; AAG27545.1; JOINED.
DR EMBL; AF261118; AAG27545.1; JOINED.
DR EMBL; AF261119; AAG27545.1; JOINED.
DR EMBL; AF261120; AAG27545.1; JOINED.
DR EMBL; AF261121; AAG27545.1; JOINED.
DR EMBL; AF261122; AAG27545.1; JOINED.
DR EMBL; AF261123; AAG27545.1; JOINED.
DR EMBL; AF261124; AAG27545.1; JOINED.
DR EMBL; AF261125; AAG27545.1; JOINED.
DR EMBL; AF261126; AAG27545.1; JOINED.
DR EMBL; AF261127; AAG27545.1; JOINED.
DR EMBL; AF261128; AAG27545.1; JOINED.
DR EMBL; AF261129; AAG27545.1; JOINED.
DR EMBL; AF261130; AAG27545.1; JOINED.
SQ SEQUENCE 1367 AA; 140454 MW; 40F259A3CD47C982 CRC64;

Query Match          59.2%; Score 573; DB 11; Length 1367;
Best Local Similarity 62.4%; Pred. No. 5.5e-49;
Matches 111; Conservative 22; Mismatches 41; Indels 4; Gaps 1;

QY 7 0PVLHLVALNTPILSGMRGIRGADFCQFOARAVGLSTGFRAFLSSRLDLYSTVRRADR 66
DB 1194 RPYLHLVALNTPVAGDIR---ADFCFOARAGLSTFRFLSSRLDLYSTVRRADR 1249
QY 67 GSVPIYNLKDENVLSPWSDLSFGSGOQLQPGARIFSPDGRDVLRRHPAPQKSVWHGSDPS 126
DB 1250 FGLPIYNLKGQVLFNWNDSJFSGDGFNTHTPIYSPDGRDVTDPSPQKVVWHGSNPH 1309
QY 127 GRRLMESYCEWTRETGTGATGQASSLSGRLLDEKAAKSNSTIVLCIENSFMTSPSK 184
DB 1310 GVRLVKDYCEAWRTTDMAYVGFASPLSTGKILDKKAYSCANRLIVLCIENSFMTDRK 1367

RESULT 9
Q9Y4W4 PRELIMINARY; PRT; 1388 AA.
AC Q9Y4W4;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE TYPE XV COLLAGEN.
GN COL15A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=94148920; PubMed=8106446;
RA Kivirikko S., Heinamaki P., Rehn M., Honkanen N., Myers J.C.,
RA Pihlajaniemi T.;
RT *Primary structure of the alpha 1 chain of human type XV collagen and
RT exon-intron organization in the 3' region of the corresponding gene.";
RL J. Biol. Chem. 269:4773-4779(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=98316357; PubMed=9651385;
RA Hagg P.M., Muona A., Lietard J., Kivirikko S., Pihlajaniemi T.;

```

RL J. Biol. Chem. 273:17824-17831(1998)

QY 7 QPVLHLVALNTPLSGGMRCIRGADFOCFQOARAVGLSGTERAFLSSRIQDIYSTVRRAD 66

RN [1]

RX MEDLINE=20196006; PubMed=10731132;

8 PVIHI VAI NTBISGCMPTBGADCECEGGA BAWCI ECTEBAEI EENI ANI VCTIUDANNOG CZ

DT 01-JAN-1998 (TREMBLEI. 05, Created)







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2001, 12:49:05 ; Search time 20.79 Seconds  
(without alignments)  
536.547 Million cell updates/sec

Title: US-09-589-777A-2

Perfect score: 968  
Sequence: 1 HTTDFQFVHLVALNTPLS.....CHNSYVLCIENSEMFSK 184

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq\_0601.\*  
1: /SID8/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID8/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID8/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID8/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID8/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID8/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID8/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID8/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID8/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID8/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID8/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID8/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID8/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID8/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID8/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID8/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID8/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID8/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID8/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID8/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID8/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	968	100.0	184	22	AAB49380
2	968	100.0	191	21	AAB28398
3	965	99.7	184	20	AAV08689
4	965	99.7	184	21	AAV70258
5	965	99.7	207	22	AAV71930
6	965	99.7	218	20	AAV08691
7	965	99.7	580	20	AAV08692
8	965	99.5	1288	18	AAW26328
9	960	99.2	184	20	AAV18409
10	960	99.2	1288	20	AAW92297
11	950	98.1	684	20	AAV25114

12	916	94.6	185	20	AAV06197
13	841	86.9	184	21	AAV70265
14	840	86.8	182	21	AAB28399
15	840	86.8	182	21	AAV94323
16	840	86.8	182	21	AAV59622
17	840	86.8	183	20	AAV08693
18	840	86.8	183	20	AAV02113
19	840	86.8	183	21	AAB30493
20	840	86.8	183	21	AAB16451
21	840	86.8	183	21	AAV90771
22	840	86.8	183	21	AAV70252
23	840	86.8	183	22	AAB49379
24	840	86.8	195	21	AAW90874
25	840	86.8	216	21	AAB30495
26	840	86.8	684	18	AAW26327
27	840	86.8	684	20	AAV25113
28	840	86.8	1301	20	AAW92296
29	840	86.8	1336	20	AAV08694
30	839	86.7	193	21	AAW90877
31	836	86.4	183	22	AAB49810
32	829	85.6	271	21	AAB08407
33	822	84.9	178	21	AAV94324
34	754	77.9	184	22	AAB49381
35	574	59.3	180	22	AAB49383
36	567	58.6	108	22	AAB49807
37	563	58.2	176	21	AAW90875
38	563	58.2	180	22	AAB49382
39	563	58.2	181	20	AAV22227
40	563	58.2	181	20	AAV18410
41	563	58.2	191	21	AAW90876
42	296	30.6	85	20	AAV22226
43	273.5	28.3	124	20	AAV26654
44	260	26.9	50	22	AAB35887
45	252	26.0	48	22	AAB49808

#### ALIGNMENTS

RESULT 1  
ID AAB49380 standard; Protein; 184 AA.  
AC AAB49380;  
XX  
DF 02-MAR-2001 (first entry)  
XX  
DE Murine endostatin SEQ ID NO: 4.  
XX  
XX Endostatin; antiangiogenic; angiogenesis; human; mouse; chicken;  
KW cancer; inflammation; angiogenesis-dependent disease.  
XX  
OS Mus musculus.  
XX  
PN WO200067771-A1.  
XX  
PD 16-NOV-2000.  
XX  
PF 02-MAY-2000; 2000MO-US12063.  
XX  
PR 06-MAY-1999; 99US-0132907.  
XX  
PR 14-JUL-1999; 99US-0353333.  
XX  
PA (BURN-) BURNHAM INST.  
XX  
PI Vuori K;  
XX  
XX WPI; 2001-040937/05.  
XX  
XX N-PSDB; AAC88290.  
XX  
XX Endostatin peptide comprising at least four endostatin amino acid  
XX residues are e.g. angiogenesis inhibitors for treating cancer and  
XX diabetic retinopathy -  
XX  
XX Murine endostatin

XX PS Disclosure: Fig 1; 146pp; English.  
XX CC The present invention provides endostatin peptides which can be used in  
CC the modulation of angiogenesis. This is useful in the treatment of  
CC cancers, inflammation, rheumatoid arthritis, chronic articular  
CC osteoarthritis, psoriasis, disorders associated with inappropriate invasion of  
CC vessels such as diabetic retinopathy, neovascular glaucoma, retinopathy  
CC of prematurity, macular degeneration, corneal graft rejection,  
CC retrolental fibroplasia, rubeosis, capillary proliferation in  
CC atherosclerotic plaques and osteoporosis. Other angiogenesis-dependent  
CC diseases include Osler-Webber syndrome, myocardial angiogenesis, plaque  
CC neovascularisation, telangiectasia, haemophilic joints and wound  
CC granulation. In addition, the peptides can be used as birth control  
CC agents.  
XX CC  
XX SO Sequence 184 AA:  
  
Query Match 100.0%; Score 968; DB 22; Length 184;  
Best Local Similarity 100.0%; Pred. No. 3.2e-108;  
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 HTHODFQPVHLVALNTPLSGMRGIRGADPOCFQOARAVGLSGTFRAFLSSRLQDLYSI 60  
DB 1 htbqdfqpvhlvalntrplsgmrgirgadfqcfqgaravglsgtfratfssrlqdllysl 60  
OY 61 VRRADRGSVPIVNLKDEVLSPSWDSLFSGSQGLQPGARIFSPDGRDYLRRHPAMPKSVW 120  
DB 61 vrradrgsvpiivnlkdevlspswdsltfsqsgqlqpgarilfsdgrdylrrhpampksw 120  
OY 121 HGSDPSGRRLMESYCEWRTETTGATGQASSLSGRLEQKASCHNSYIVLCIENSFMT 180  
DB 121 hgdspsgrrlmesycewrtettgatgqassllsgrlleqkaaschnsyivlcienf 180  
OY 181 SFSK 184  
DB 181 sfsk 184  
  
RESULT 2  
ID AAB28398 standard; Protein; 191 AA.  
XX AC AAB28398;  
XX DT 19-FEB-2001 (first entry)  
XX DE Murine endostatin.  
XX KW Murine; endostatin; cytostatic; antiproliferative;  
KM vascular endothelial growth factor; VEGF; antibody; VEGF2 receptor;  
XX cancer; vascularised solid tumour.  
XX OS Mus sp.  
XX PN MO200064946-A2.  
XX PD 02-NOV-2000.  
XX PF 28-APR-2000; 2000MO-US11367.  
XX PR 28-APR-1999; 99US-0131432.  
XX PA (TEXA ) UNIV TEXAS SYSTEM.  
XX PI Thorpe PE, Brekken RA;  
XX WPI; 2000-687317/67.  
XX DR N-PSDB; AAC67777.  
XX PT Immunogenic composition for the treatment and diagnosis of cancer  
PT comprises an anti-VEGF (vascular endothelial growth factor) antibody

PT binding the same epitope as the monoclonal antibody ATCC PTA 1595 -  
XX XX  
XX Example 10; Page 290-291; 298pp; English.  
XX CC The present invention relates to anti-Vascular Endothelial Growth Factor  
CC (VEGF) antibodies that bind to the same epitope as the monoclonal  
CC antibody ATCC PTA 1595 and which significantly inhibit VEGF binding to  
CC the VEGF receptor VEGFR2, without inhibiting VEGF binding to the VEGF  
CC receptor VEGFR1. The present sequence is murine endostatin. Endostatin  
CC may be conjugated onto the anti-VEGF antibodies of the present invention.  
CC The anti-VEGF antibodies of the present invention are useful for the  
CC treatment and diagnosis of cancer, especially vascularised solid tumours.  
XX CC  
XX SO Sequence 191 AA:  
  
Query Match 100.0%; Score 968; DB 21; Length 191;  
Best Local Similarity 100.0%; Pred. No. 3.3e-108;  
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 HTHODFQPVHLVALNTPLSGMRGIRGADPOCFQOARAVGLSGTFRAFLSSRLQDLYSI 60  
DB 8 htbqdfqpvhlvalntrplsgmrgirgadfqcfqgaravglsgtfratfssrlqdllysl 67  
OY 61 VRRADRGSVPIVNLKDEVLSPSWDSLFSGSQGLQPGARIFSPDGRDYLRRHPAMPKSVW 120  
DB 68 vrradrgsvpiivnlkdevlspswdsltfsqsgqlqpgarilfsdgrdylrrhpampksw 127  
OY 121 HGSDPSGRRLMESYCEWRTETTGATGQASSLSGRLEQKASCHNSYIVLCIENSFMT 180  
DB 128 hgdspsgrrlmesycewrtettgatgqassllsgrlleqkaaschnsyivlcienf 187  
OY 181 SFSK 184  
DB 188 sfsk 191  
  
RESULT 3  
ID AAY08689 standard; Protein; 184 AA.  
XX AC AAY08689;  
XX DT 10-AUG-1999 (first entry)  
XX DE Murine endostatin protein fragment.  
XX KW Plasmidogen; murine; angiotensin; endostatin; gene therapy; vector;  
KM anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;  
XX tumour growth; solid tumour; diabetic retinopathy; retina.  
XX OS Mus sp.  
XX PN MO9926480-A1.  
XX PD 03-JUN-1999.  
XX PF 20-NOV-1998; 98MO-US24950.  
XX PR 20-NOV-1997; 97US-0975424.  
XX PA (GENE-) GENETIX PHARM INC.  
XX PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
XX PI Bachelot T, Leboulch P, Pawluc R;  
XX WPI; 1999-357696/30.  
XX DR N-PSDB; AAX77715.  
XX PT Anti-angiogenic gene therapy vectors  
XX PS Disclosure: Fig 6; 83pp; English.  
XX XX



CC This invention describes a novel viral gene therapy vector comprising a  
CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen  
CC from human or murine angiostatin, human or murine endostatin and  
CC angiogenesis-inhibiting fusions and fragments, where the viral vector is  
CC sufficiently attenuated for use in human gene therapy. The products of  
CC the invention have anti-angiogenic, cytostatic, anti-diabetic and  
CC ophthalmological activity. The vector is used in gene therapy for  
CC inhibiting tumour growth in humans harbouring a solid tumour. The vector  
CC expresses an anti-angiogenic polypeptide. An additional use comprises  
CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide  
CC inhibits angiogenesis in the vicinity of the retina. The vector is  
CC administered to cells ex vivo and then administered to the patient.

XX Sequence 184 AA:

Query Match 99.7%; Score 965; DB 20; Length 184;  
Best Local Similarity 99.5%; Pred. No. 7.3e-108;  
Matches 183; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 HTHQDFQPVHLVALNTPLSGMRGIRGADFOCFQOARAVGLSGTFRALFSLRLOLXSI 60  
DB 1 hthqdfqpvhlvalntplsgmrgirgadfgcfqgaravglsgtfralfslrlqlysl 60  
OY 61 VRRADRGSPVPIYNLKDENVLSFSGSQGLOPGARIFSGDGVLRHPAMPQKSYW 120  
DB 61 vrradrgsvpiynlkdevlspswdsifsgsqgvgqgarifsgdgvlrhpampqksw 120  
OY 121 HGSDPSGRRLMESYCTWRTETTGATGOASSLSLGRLEOKAASCHNSYIVLCIENSFMT 180  
DB 121 hgsdpsgrlmesyctwrtettgatgqassllsgrlleqkaaschmsyivlcienfnt 180  
OY 181 SFSK 184  
DB 181 sfsk 184

RESULT 4

AAV70258  
ID AAV70258 standard; Protein: 184 AA.

XX AAV70258;

DE 06-JUN-2000 (first entry)

XX Murine angiogenesis inhibitor, endostatin.

KM Murine; immunoglobulin Fc fragment; endostatin; immunofusin;  
KM angiogenesis; inhibitor; cytostatic; antirheumatoid; antiarthritic;  
KM antipsoriatic; antidiabetic; ophthalmological; immunosuppressant;  
KM vasotrophic; vulnery; treatment; antiarteriosclerosis; tumour;  
KM metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;  
KM ocular angiogenic disease; diabetic retinopathy; macular degeneration;  
KM myocardial angiogenesis; plaque neovascularisation; telanglectasia;  
KM wound granulation; keloid scar; gene therapy.

XX Mus musculus.

OS WO200011033-A2.

PN 02-MAR-2000.

XX 25-AUG-1999; 99WO-US19329.

XX 25-AUG-1998; 98US-0097883.

XX (LEXI-) LEXINGEN PHARM CORP.

XX Le K, Li Y, Gillies SD;

XX WPI, 2000-237616/20.

DR N-PSDB; AA251299.

PT Novel fusion protein of angiostatin or endostatin and an immunoglobulin  
PT FC region, useful for treating conditions mediated by angiogenesis,  
PT such as rheumatoid arthritis, tumors and macular degeneration -  
XX Example 5; Pages 48-49; 68pp; English.

XX The patent discloses a DNA molecule encoding a fusion protein comprising  
XX a signal sequence, an immunoglobulin Fc region, and an angiogenesis  
XX inhibitor selected from angiostatin, endostatin, a plasminogen fragment  
XX having angiostatin activity, a collagen XVIII fragment having endostatin  
XX activity, or combinations of them. The fusion protein (immunofusin) is  
XX used to inhibit angiogenesis and to treat diseases or conditions mediated  
XX by angiogenesis. Conditions that may be treated include solid tumours,  
XX blood born tumours, tumour metastasis, benign tumours including  
XX haemangiomas, acoustic neuromas, neurofibromas, trachomas and pyrogenic  
XX granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases  
XX e.g. diabetic retinopathy, retinopathy of prematurity, macular  
XX degeneration, corneal graft rejection, neovascular glaucoma, retrolental  
XX plaque neovascularisation, telanglectasia, haemophilic joints,  
XX angiofibroma, wound granulation, and excessive or abnormal stimulation of  
XX endothelial cells, intestinal cells, atherosclerosis, sclerodermal and  
XX hypertrophic scars, i.e. keloid scars. The DNA constructs may be used  
XX in gene therapy. The present sequence is a murine  
XX immunoglobulin Fc fragment.

XX Sequence 184 AA:

Query Match 99.7%; Score 965; DB 21; Length 184;  
Best Local Similarity 99.5%; Pred. No. 7.3e-108;  
Matches 183; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 HTHQDFQPVHLVALNTPLSGMRGIRGADFOCFQOARAVGLSGTFRALFSLRLOLXSI 60  
DB 1 hthqdfqpvhlvalntplsgmrgirgadfgcfqgaravglsgtfralfslrlqlysl 60  
OY 61 VRRADRGSPVPIYNLKDENVLSFSGSQGLOPGARIFSGDGVLRHPAMPQKSYW 120  
DB 61 vrradrgsvpiynlkdevlspswdsifsgsqgvgqgarifsgdgvlrhpampqksw 120  
OY 121 HGSDPSGRRLMESYCTWRTETTGATGOASSLSLGRLEOKAASCHNSYIVLCIENSFMT 180  
DB 121 hgsdpsgrlmesyctwrtettgatgqassllsgrlleqkaaschmsyivlcienfnt 180  
OY 181 SFSK 184  
DB 181 sfsk 184

RESULT 5

AAV71930  
ID AAV71930 standard; Protein: 207 AA.

XX AAV71930;

DE 10-MAY-2001 (first entry)

XX Murine endostatin attached to Ig-kappa leader sequence.

KM Mouse; endostatin; antitumour; cytostatic; antiarthritic; antipsoriatic;  
KM antidiabetic; ophthalmological; gene therapy; angiogenic inhibitor;  
KM adenoviral vector; diabetic retinopathy; cardiovascular disease;  
KM arthritis; psoriasis; cerebral oedema; intravascular coagulopathy;  
KM lymphoma; leukaemia; immunoglobulin; Ig; Ig-kappa.

XX Mus sp.

XX WO2000112830-A1.

XX 22-FEB-2001.

PF 11-AUG-2000; 2000MO-EP07865.  
 XX  
 PR 13-AUG-1999; 99US-0373938.  
 XX  
 PA (NOVS ) NOVARTIS AG.  
 PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.  
 XX  
 PI Hallenbeck PL, Chen CT;  
 XX  
 DR WPI: 2001-202871/20.  
 DR N-PSDB; AAF60336.  
 PT Adenoviral vector for treating tumors and disorders associated with  
 PT angiogenesis, such as cancer, arthritis, and psoriasis, comprises a DNA  
 PT sequence encoding an angiogenic inhibitor, particularly endostatin  
 XX  
 PS Example 1; Fig 1B; 59pp; English.  
 CC The nucleotide sequence encoding this protein was used in the  
 CC construction of an adenoviral vector which includes a DNA sequence  
 CC encoding endostatin. The adenoviral vector is useful for expressing  
 CC endostatin in a mammalian cell such as an A549 or Hep3B cell. It is  
 CC useful for treating other diseases and disorders associated with  
 CC angiogenesis, such as neovascular diseases of the eye, including diabetic  
 CC retinopathy, cardiovascular disease, arthritis, psoriasis, cerebral  
 CC edema and intravascular coagulopathy (Kasabach-Merritt syndrome). The  
 CC vector inhibits, prevents or destroys the growth of tumors by  
 CC preventing the formation of blood vessels in tumors, such as lymphoma  
 CC and leukaemia.  
 XX  
 SQ Sequence 207 AA:

Query Match 99.7%; Score 965; DB 22; Length 207;  
 Best Local Similarity 99.5%; Pred. No. 8.6e-108;  
 Matches 183; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HTHDQFQVPLHLVALNTPLSGMRGIRGADFCFOQARAVGLSGTFRAFLSSRLQDLYSI 60  
 DB htbqdfqpvhlvalntrpplsgmrgirgadfcfqgaravglsgtfrfaflssrlqdlysl 83  
 QY 61 VRRADRGSVPIVNLKDEVLSPSMDSLFSGSGQLQPCARIFSPDGRVYLHRPMPQKSVW 120  
 DB vrradrgsvpiivnlkdevlspswdsltsgsgqqlqpcarilfsdpgrvylhrmpmpqksw 143  
 QY 121 HGSPPSGRRLMESYCEFTWRETTGATGQASSLSGRLLEQKAASCHNSYIVLCIENSFMT 180  
 DB hgsppsgrrlmesycefwtrettgatgqasslsgrlleqkaaschnsyivlcienfmc 203  
 QY 181 SFSK 184  
 DB 204 sfsk 207

RESULT 6  
 AAY08691  
 ID AAY08691 standard; Protein: 218 AA.  
 XX  
 AC AAY08691;  
 XX  
 DT 10-AUG-1999 (first entry)  
 XX  
 DE Murine gene therapy peptide construct SP-Flag-Endo.  
 XX  
 KW Plasmidogen: murine; angiotatin; endostatin; gene therapy; vector;  
 KW anti-angiogenic; attenuation; cytosstatic; anti-diabetic; ophthalmology;  
 KW tumour growth; solid tumour; diabetic retinopathy; retina; construct.  
 XX  
 OS Mus sp.  
 OS Synthetic.  
 XX  
 PN W09926480-A1.  
 XX

PD 03-JUN-1999.  
 XX  
 PF 20-NOV-1998; 98MO-US24950.  
 XX  
 PR 20-NOV-1997; 97US-0975424.  
 XX  
 PA (GENE-) GENETIX PHARM INC.  
 PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
 XX  
 PI Bachelot T, Leboulch P, Pawliuk RJ;  
 XX  
 DR WPI: 1999-357696/30.  
 DR N-PSDB; AAX77717.  
 PT Anti-angiogenic gene therapy vectors  
 XX  
 PS Example 1; Page 69; 83pp; English.  
 CC This invention describes a novel viral gene therapy vector comprising a  
 CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen  
 CC from human or murine angiotatin, human or murine endostatin and  
 CC angiogenesis-inhibiting fusions and fragments, where the viral vector is  
 CC sufficiently attenuated for use in human gene therapy. The products of  
 CC the invention have anti-angiogenic, cytosstatic, anti-diabetic and  
 CC ophthalmological activity. The vector is used in gene therapy for  
 CC inhibiting tumor growth in humans harbouring a solid tumour. The vector  
 CC expresses an anti-angiogenic polypeptide. An additional use comprises  
 CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide  
 CC inhibits angiogenesis in the vicinity of the retina. The vector is  
 CC administered to cells ex vivo and then administered to the patient.  
 XX  
 SQ Sequence 218 AA:

Query Match 99.7%; Score 965; DB 20; Length 218;  
 Best Local Similarity 99.5%; Pred. No. 9.3e-108;  
 Matches 183; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HTHDQFQVPLHLVALNTPLSGMRGIRGADFCFOQARAVGLSGTFRAFLSSRLQDLYSI 60  
 DB htbqdfqpvhlvalntrpplsgmrgirgadfcfqgaravglsgtfrfaflssrlqdlysl 94  
 QY 61 VRRADRGSVPIVNLKDEVLSPSMDSLFSGSGQLQPCARIFSPDGRVYLHRPMPQKSVW 120  
 DB vrradrgsvpiivnlkdevlspswdsltsgsgqqlqpcarilfsdpgrvylhrmpmpqksw 154  
 QY 121 HGSPPSGRRLMESYCEFTWRETTGATGQASSLSGRLLEQKAASCHNSYIVLCIENSFMT 180  
 DB hgsppsgrrlmesycefwtrettgatgqasslsgrlleqkaaschnsyivlcienfmc 214  
 QY 181 SFSK 184  
 DB 215 sfsk 218

RESULT 7  
 AAY08692  
 ID AAY08692 standard; Protein: 580 AA.  
 XX  
 AC AAY08692;  
 XX  
 DT 10-AUG-1999 (first entry)  
 XX  
 DE Murine gene therapy peptide construct SP-K1-K2-K3-K4-Flag-Endo.  
 XX  
 KW Plasmidogen: murine; angiotatin; endostatin; gene therapy; vector;  
 KW anti-angiogenic; attenuation; cytosstatic; anti-diabetic; ophthalmology;  
 KW tumour growth; solid tumour; diabetic retinopathy; retina; construct.  
 XX  
 OS Mus sp.  
 OS Synthetic.  
 XX  
 PN W09926480-A1.  
 XX

FT	Peptide	309..314
FT	/label= GXXGX 'Y' _motif	
FT	Peptide	315..320
FT	/label= GXXGX 'Y' _motif	
FT	Peptide	321..326
FT	/label= GXXGX 'Y' _motif	
FT	Peptide	337..342
FT	/label= GXXGX 'Y' _motif	
FT	Peptide	343..348
FT	/label= GXXGX 'Y' _motif	
FT	Peptide	349..354
FT	/label= GXXGX 'Y' _motif	
FT	Peptide	355..360
FT	/label= GXXGX 'Y' _motif	
FT	Peptide	361..366
FT	/label= GXXGX 'Y' _motif	
FT	Peptide	367..372
FT	/label= GXXGX 'Y' _motif	
FT	Peptide	373..378
FT	/label= GXXGX 'Y' _motif	
FT	Peptide	379..384
FT	/label= GXXGX 'Y' _motif	
FT	Peptide	385..390
FT	/label= GXXGX 'Y' _motif	
FT	Peptide	396..401
FT	/label= GXXGX 'Y' _motif	
FT	Peptide	402..407
FT	/label= GXXGX 'Y' _motif	
FT	Peptide	433..440
FT	/label= GXXGX 'Y' _motif	
FT	Peptide	441..446
FT	/label= GXXGX 'Y' _motif	
FT	Peptide	447..452

FT	Peptide	
FT	reptide	/label=GXGX'Y', motif=453..456
FT		/label=GXGX'Y', motif=459..464
FT		/label=GXGX'Y', motif=459..464

Protein	Peptide	Label	Score
FT	Peptide	/label= GYXGX/Y',_motif	476.481
FT	Peptide	/label= GYXGX/Y',_motif	487.487
FT	Peptide	/label= GYXGX/Y',_motif	488.493
FT	Peptide	/label= GYXGX/Y',_motif	494.499
FT	Peptide	/label= GYXGX/Y',_motif	500.505
FT	Peptide	/label= GYXGX/Y',_motif	506.511
FT	Peptide	/label= GYXGX/Y',_motif	512.517
FT	Peptide	/label= GYXGX/Y',_motif	

FT	Peptide	/label= GYXGX 'Y' _motif	524..529
FT	Peptide	/label= GYXGX 'Y' _motif	530..535
FT	Peptide	/label= GYXGX 'Y' _motif	536..541
FT	Peptide	/label= GYXGX 'Y' _motif	542..547
FT	Peptide	/label= GYXGX 'Y' _motif	548..553
FT	Peptide	/label= GYXGX 'Y' _motif	560..585
FT	Peptide	/label= GYXGX 'Y' _motif	586..591
FT	Peptide	/label= GYXGX 'Y' _motif	592..597
FT	Peptide	/label= GYXGX 'Y' _motif	598..603



KM Oslar-Webber Syndrome; myocardial angiogenesis; angiofibroma; cancer;  
 KM plaque neovascularisation; telangiectasia; atherosclerosis; scleroderma;  
 KM dialysis graft vascular access stenosis; renal cancer; therapy.  
 OS Mus sp.  
 XX WO929855-A1.  
 PN 17-JUN-1999.  
 XX  
 XX 08-DEC-1998: 98WO-US26057.  
 PF  
 XX 16-NOV-1998: 98US-0108536.  
 PR 08-DEC-1997: 97US-0067888.  
 PR 22-APR-1998: 98US-0082663.  
 XX  
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
 PI Sukhathave VP;  
 XX  
 DR WPI: 1999-385604/32.  
 DR N-PSDB; AAX79949.  
 XX  
 PT Mutant endostatin having anti-angiogenic activity  
 PS Claim 31; Fig 2; 105pp; English.  
 XX  
 CC This sequence is the mouse endostatin. The invention relates to a  
 CC the mutant endostatin (EM), which has anti-angiogenic activity, and is  
 CC designated EM1. Compositions comprising EM1 or fusion proteins comprising  
 CC EM1, are useful for treating diseases characterised by angiogenic  
 CC activity, such as angiogenesis-dependent cancers, benign tumours,  
 CC rheumatoid arthritis, psoriasis, ocular angiogenesis, Oslar-Webber  
 CC Syndrome, myocardial angiogenesis, plaque neovascularisation,  
 CC telangiectasia, haemophilic joints, angiofibroma, wound granulation,  
 CC intestinal adhesions, atherosclerosis, scleroderma, hypertrophic scars,  
 CC cat scratch disease, Helicobacter pylori ulcers, dialysis graft vascular  
 CC access stenosis, contraception and obesity. In particular, the diseases  
 CC treatable by EM1 comprise cancer, especially renal cancer. The methods  
 CC provide a means for introducing EM1 into mammalian cells via gene  
 CC therapy, for production of EM1 via recombinant means, as well as  
 CC recombinant production of the EM1 protein. EM1 performs as well or better  
 CC than whole endostatin. Use of EM1 is advantageous for treatment of  
 CC angiogenic diseases in that increasingly smaller peptides are more potent  
 CC on a weight basis, and may be able to better penetrate tissues.  
 CC  
 XX Sequence 184 AA;  
 SQ  
 Query Match 99.5%; Score 963; DB 20; Length 184;  
 Best Local Similarity 99.5%; Pred. No. 1.3e-107;  
 Matches 183; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 HTHOFOPLVHLVANTPLSGMGRIGADPOCARAVGSGFRFLSLRLDLYSI 60  
 DB 1 hthgdfgpyhlhvalncplsggmrglrgadqctqgaravgsiftrflssrlldlysi 60  
 QY 61 VRADRGSVPIVNLKDEVLSPSWDSLFGSGGOLQPGARIFSPGDRDYLRRHPAMPKSW 120  
 DB 61 vrradrgsvplvnlkdevlspswdsifsgsgqlpgarilfsifgrdylrrhpampqksv 120  
 QY 121 HGSDDSGRRIMSEYCEWRTETTGATGQASSLSGRLLERKAAACHNSYIVLCIENSFMT 180  
 DB 121 hgsddsgrrimesycewrtetttgatgqassllsgrllleqkaaschnsyivlcienfnt 180  
 QY 181 SFSK 184  
 DB 181 sfsk 184  
 RESULT 10  
 AAW92297 X  
 ID AAW92297 standard; peptide: 1288 AA.

XX  
 AC AAW92297;  
 XX  
 DT 28-APR-1999 (first entry)  
 XX  
 DE Mouse alpha-1 (XVIII) collagen chain common sequence MO18(common)28.  
 XX  
 KM Human; type XVIII collagen; liver disease; cirrhosis; detection;  
 KM hepatocellular carcinoma; diagnosis.  
 XX  
 OS Mus sp.  
 XX  
 PN WO9856399-A1.  
 XX  
 PD 17-DEC-1998.  
 XX  
 PF 12-JUN-1998: 98WO-US12327.  
 PR 12-JUN-1997: 97US-0049369.  
 XX  
 PA (FIR-) ACAD FINLAND.  
 PA (FIR-) FIBROGEN INC.  
 PA (INRM) INST NAT SANTE & RECH MEDICALE.  
 PI Clement B, Pihlajaniemi T, Rehn M;  
 XX  
 DR WPI: 1999-070292/06.  
 XX  
 PT Diagnosis and monitoring of liver disease by measuring collagen type  
 PT XVIII levels - with elevated levels indicative of disease,  
 PT especially cirrhosis or hepatocellular carcinoma  
 XX  
 PS Example 6; Fig 8; 56pp; English.  
 XX  
 CC A method has been developed for the detecting liver disease. The method  
 CC comprises: (a) reacting a patient sample with antibodies (Ab) specific  
 CC for collagen type XVIII (Col18); (b) measuring the amount of Ab-antigen  
 CC complex (c) formed as indicator of the amount of Col18 present; (c)  
 CC similar analysis of a non-diseased control; and (d) comparing the  
 CC amounts of Col18 in the two samples to detect presence or progression of  
 CC disease. Elevated levels of Col18 are: (i) indicative of disease,  
 CC specifically cirrhosis; and (ii) predictive of the prognosis of disease,  
 CC Col18 mRNA levels and tumour size and necrosis, and survival times are  
 CC significantly higher in patients with higher Col18 levels). The method  
 CC provides non-invasive, early and accurate diagnosis of liver disease.  
 CC The present sequence represents the sequence common to mouse alpha-1  
 CC (XVIII) collagen chain from the present invention.  
 CC  
 XX Sequence 1288 AA;  
 SQ  
 Query Match 99.2%; Score 960; DB 20; Length 1288;  
 Best Local Similarity 99.5%; Pred. No. 5.2e-106;  
 Matches 182; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HTHOFOPLVHLVANTPLSGMGRIGADPOCARAVGSGFRFLSLRLDLYSI 60  
 DB 1106 hthgdfgpyhlhvalncplsggmrglrgadqctqgaravgsiftrflssrlldlysi 1165  
 QY 61 VRADRGSVPIVNLKDEVLSPSWDSLFGSGGOLQPGARIFSPGDRDYLRRHPAMPKSW 120  
 DB 1166 vrradrgsvplvnlkdevlspswdsifsgsgqlpgarilfsifgrdylrrhpampqksv 1225  
 QY 121 HGSDDSGRRIMSEYCEWRTETTGATGQASSLSGRLLERKAAACHNSYIVLCIENSFMT 180  
 DB 1226 hgsddsgrrimesycewrtetttgatgqassllsgrllleqkaaschnsyivlcienfnt 1285  
 QY 181 SFS 183  
 DB 181 sfs 183  
 DB 1286 sfs 1288

	Query Match	Best Local Similarity	Matches 180;	Conservative	Score 950;	Ped. No. 3.2e-105;	DB 20;	Length 664;
QY	1	HTHQDQFVPLVILVATLMPISGSGMRGIRGADFOCFQOARAVGLSGTFRAFLSRLQDLYSI	60					
DB	502	htqdfqfpvlhvalhltcrlpsgmrglrtpaafqcfqfqravrglsqftralfssrlqdyss	561					
QY	61	VRRADGSGVPTVLAKQDEVLSPPSNDLSLFSGSQGGLQGCARLFSPDGRDLRHPMPKPSVW	120					
DB	562	vrradqgsrvplvllkdevlspwdsflfsgsgqvgvqparlfstcdgdrvltnhpwqpksvw	621					
QY	121	HGSDPSGRRLMESEYCEFWRTETTGANGQASLLSGRLLEOKAASCHNSYIVLCEIENSPMT	180					
DB	622	hgadpsgrtrlmesyccetwrtetlgatqgaassllsgrllleqkaaschmsyivlciensfmlt	661					

QY	181 S 181
DB	682 S 682
RESULT 12	
ID	AAV06197
XX	AAV06197 standard; Protein: 185 AA.
AC	AAV06197;
DT	16-AUG-1999 (first entry)
XX	
DE	Anti-angiogenic endostatin peptide.
XX	
KM	Anti-angiogenic; carrier:DNA complex; tumour; gene therapy; human; endostatin; melanoma; lung cancer; colon cancer; brain cancer; breast cancer.
KW	
OS	Homo sapiens.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 36 /note= "encoded by CAG"
FT	Misc-difference 37 /note= "encoded by CAA"
FT	Misc-difference 39 /note= "deduced sequence from nucleotide sequence has an Ala residue between residues 39 and 40 of this sequence"
FT	Misc-difference 76 /note= "encoded by AAG"
FT	Misc-difference 118 /note= "encoded by AAG"
FT	Misc-difference 162 /note= "encoded by AAA"
FT	Misc-difference 168 /note= "encoded by AAC"
FT	Misc-difference 185 /note= "encoded by AAA"
FT	
PN	EP921193-A1.
XX	
PD	09-JUN-1999.
XX	
PF	07-JAN-1998; 98EP-0100135.
XX	
PR	05-DEC-1997; 97US-0985526.
XX	
PA	(MIXS/) MIXSON A J.
PI	Mixson AJ;
XX	
DR	WPI: 1999-315406/27.
DR	N-PSDB; AAX58740.
XX	
PT	Inhibition of growth of solid tumors
XX	
PS	Disclosure: Page 38; 46pp; English.
XX	
CC	The present sequence represents an anti-angiogenic endostatin peptide. The invention provides a carrier:DNA complex that comprises DNA (see AAX58725-42) encoding an anti-angiogenic protein or peptide, such as the present sequence, the complex being deliverable to the site of a tumour in vivo, and additionally comprises regulatory elements for expressing the anti-angiogenic DNA in a tumour or tumour vasculature. The complex may also include DNA encoding a tumour suppressor protein, especially p53. The carrier is a liposome, cationic polymer, micelle, microsphere, virus, viral component, or a combination of these, and administration is by intravenous or intratumoral injection. The complexes are useful in gene therapy for treatment of tumour growth. The types of tumors which may be treated include solid tumors such as melanomas and

CC tumors in the lung, colon, brain and breast.  
 XX Sequence 185 AA;

Query Match 94.6%; Score 916; DB 20; Length 185;  
 Best Local Similarity 95.1%; Pred. No. 5.7e-102;  
 Matches 176; Conservative 5; Mismatches 2; Indels 2; Gaps 2;

QY 1 HTHQDFOPVHLVLTPLSGMRGIRGADPOCFQARAVGSGTFRALFSSRLQDLXST 60  
 Db 2 hthqdfqpylhvltplsgmrgirgadfqrmar-vqtsqtralfssrlqdlxst 60  
 QY 61 VRADRGSPVIV-NLKDEVLSFSDSLFSGSQGQLQPCARIFSPGKRDVLRHPAPQKSV 119  
 Db 61 vrradrgspvivrdevlfspswdsllfsgsqqlpqrarlfsgdrvrlrhpapqrsv 120  
 QY 120 WHGSDSGRRLMESYCEWTETTGATGQASSLSGRLLEQKASCHNSYIVLCIENSFM 179  
 Db 121 whgsdpsgrllmesyctwtetgtgqassllsgrlleqraaschdsyivlcienfsm 180  
 QY 180 TSFSK 184  
 Db 181 tsfsr 185

RESULT 13  
 AAY70265  
 ID AAY70265 standard; Protein; 184 AA.  
 AC AAY70265;

DT 06-JUN-2000 (first entry)  
 XX  
 DE Canine angiogenesis inhibitor, endostatin.

XX Canine: immunoglobulin Fc fragment; endostatin; immunofusin;  
 KW angiogenesis; inhibitor; cytostatic; antirheumatoid; antiarthritis;  
 KW antiproliferative; antidiabetic; ophthalmological; immunosuppressant;  
 KW vasotropic; vulnary; treatment; antiarteriosclerosis; tumour;  
 KW metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;  
 KW ocular angiogenic disease; diabetic retinopathy; macular degeneration;  
 KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;  
 KW wound granulation; keloid scar; gene therapy.

XX Cantis familiaris.

XX WO200011033-A2.

XX 02-MAR-2000.

XX 25-AUG-1999; 99WO-US19329.

XX 25-AUG-1998; 98US-0097883.

XX (LEXI-) LEXINGEN PHARM CORP.

XX Lo K, Li Y, Gillies SD;

XX WPI; 2000-237616/20.

XX N-PSDB; AAZ51309.

XX Novel fusion protein of angiotensin or endostatin and an immunoglobulin  
 PT FC region, useful for treating conditions mediated by angiogenesis,  
 XX such as rheumatoid arthritis, tumors and macular degeneration -

XX Example 8: Pages 59-60; 68pp; English.

XX The patent discloses a DNA molecule encoding a fusion protein comprising  
 CC a signal sequence, an immunoglobulin Fc region, and an angiogenesis  
 CC inhibitor selected from angiotensin, endostatin, a plasminogen fragment  
 CC having angiotensin activity, a collagen XVIII fragment having endostatin  
 CC activity, or combinations of them. The fusion protein (immunofusin) is

CC used to inhibit angiogenesis and to treat diseases or conditions mediated  
 CC by angiogenesis. Conditions that may be treated include solid tumours,  
 CC blood born tumours, tumour metastasis, benign tumours including  
 CC haemangiomas, acoustic neuromas, neurofibromas, trachomas and pyrogenic  
 CC granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases  
 CC e.g. diabetic retinopathy, retinopathy of prematurity, macular  
 CC degeneration, corneal graft rejection, neovascular glaucoma, retrolental  
 CC plaque neovascularisation, telangiectasia, haemophilic joints,  
 CC angiodiroma, wound granulation, and excessive or abnormal stimulation of  
 CC endothelial cells, intestinal cells, atherosclerosis, scleroderma and  
 CC hypertrophic scars, i.e. keloid scars. The DNA constructs may be used  
 CC in gene therapy. The present sequence is a canine  
 CC endostatin used in the construction of immunofusin containing canine  
 CC immunoglobulin Fc fragment.

XX Sequence 184 AA;

Query Match 86.9%; Score 841; DB 21; Length 184;  
 Best Local Similarity 84.2%; Pred. No. 6e-93;  
 Matches 155; Conservative 17; Mismatches 12; Indels 0; Gaps 0;

QY 1 HTHQDFOPVHLVLTPLSGMRGIRGADPOCFQARAVGSGTFRALFSSRLQDLXST 60  
 Db 1 hthqdfqpylhvltplsgmrgirgadfqrmar-vqtsqtralfssrlqdlxst 60  
 QY 61 VRADRGSPVIV-NLKDEVLSFSDSLFSGSQGQLQPCARIFSPGKRDVLRHPAPQKSV 120  
 Db 61 vrradrgspvivrdevlfspswdsllfsgsqqlpqrarlfsgdrvrlrhpapqrsv 120  
 QY 121 HGSDPSGRRLMESYCEWTETTGATGQASSLSGRLLEQKASCHNSYIVLCIENSFM 180  
 Db 121 hgsdpsgrllmesyctwtetgtgqassllsgrlleqraaschdsyivlcienfsm 180  
 QY 181 SFSK 184  
 Db 181 sfsk 184

RESULT 14  
 AAB28399  
 ID AAB28399 standard; Protein; 182 AA.

AC AAB28399;

DT 19-FEB-2001 (first entry)

XX Human endostatin.

XX Human: endostatin; cytostatic; antiproliferative;  
 KW vascular endothelial growth factor; VEGF; antibody; VEGF2 receptor;  
 KW cancer; vascularised solid tumour.

XX Homo sapiens.

XX WO200064946-A2.

XX 02-NOV-2000.

XX 28-APR-2000; 2000WO-US11367.

XX 28-APR-1999; 99US-0131432.

XX (TEXA ) UNIV TEXAS SYSTEM.

XX Thorpe PE, Brecken RA;

XX WPI; 2000-687317/67.

XX Immunogenic composition for the treatment and diagnosis of cancer  
 PT comprises an anti-VEGF (vascular endothelial growth factor) antibody  
 PT binding the same epitope as the monoclonal antibody ATCC PTA 1595 -

XX PS Example 10: Page 291-292: 298pp: English.

XX CC The present invention relates to anti-Vascular Endothelial Growth Factor (VEGF) antibodies that bind to the same epitope as the monoclonal antibody AICC PTA 1595 and which significantly inhibit VEGF binding to the VEGF receptor VEGFR2, without inhibiting VEGF binding to the VEGF receptor VEGFR1. The present sequence is human endostatin. Endostatin may be conjugated onto the anti-VEGF antibodies of the present invention. CC CC The anti-VEGF antibodies of the present invention are useful for the treatment and diagnosis of cancer, especially vascularised solid tumours.

XX SQ Sequence 182 AA:

QY Query Match 86.8%; Score 840; DB 21; Length 182;  
Best Local Similarity 85.6%; Pred. No. 7.7e-93;  
Matches 155; Conservative 15; Mismatches 11; Indels 0; Gaps 0;

DB 1 HTHQDFQPVHLVLTMPISGMRGIRGADFOCFQOARAVGLSGFRFLSRLODLYSI 60  
1 hshrdtqpvhlhvalnspisgmrgirgadfcfgqaravglagtlfrrflssrlqdllysi 60

QY 61 VRRADRGSVPIYNLKDEVSLFSGSGQLQPGARIFSGRDVLRHPAMPQKSW 120  
61 vrradtraavpivnlkdelifswsaelfsgeegplkpgariffsdgkdvlrhptwpqksw 120

DB 121 HGSDPSGRRLMESYCEWTRETTGATGQASSLLSGRLLEOKASCHNSYIVLCIENSFMT 180  
121 hgsdpsgrlrlesyctwrtettgatgqassllsgrlleokaschnsyivlcienfnt 180

QY 181 S 181  
181 a 181

DB 181 a 181

RESULT 15  
ID AAY94323 standard; Protein: 182 AA.  
AC AAY94323;  
DT 11-AUG-2000 (first entry)  
DE Human endostatin protein.  
XX Human: endothelial cell proliferation inhibitor; collagen XVIII;  
KW angiogenesis inhibitor; anti-tumour; cytostatic; antiproliferative;  
KW vasotrophic; dermatological; ophthalmological; vulnary;  
KW antiarteriosclerotic; antidiabetic; haemostatic; contraceptive;  
KW ocular angiogenic disease; atherosclerosis; scleroderma;  
KW myocardial angiogenesis; telangiectasia; angiodysplasia;  
KW wound granulation.  
XX Homo sapiens.  
OS WO200026368-A2.  
FN 11-MAY-2000.  
PD 01-NOV-1999: 99WO-US25605.  
XX 30-OCT-1998: 98US-0106343.  
PR 20-MAY-1999: 99US-0315689.  
XX (CHIL-) CHILDRENS MEDICAL CENT.  
PA O'Reilly MS, Folkman MJ;  
XX WPI: 2000-365617/31.  
DR N-PSDB, AAA27004.  
XX Novel endostatin capable of inhibiting endothelial cell proliferation

PT and angiogenesis, useful for treating angiogenesis-dependent cancers  
PT and as birth control agents

XX PS Claim 2: Page 38: 68pp: English.

XX CC The present sequence is an endostatin protein which is the carboxy terminal protein of human collagen XVIII. Recombinant mouse endostatin (20 mg/kg) was administered subcutaneously to mice implanted with Lewis lung carcinomas. There was tumour mass regression non-detectable levels after 12 days of therapy due to the angiogenesis inhibitory activity of endostatin. Thus the protein is useful for treatment of angiogenesis-dependent cancers. The polynucleotide and diagnosis of tumours, ocular endostatin are useful for treating and diagnosis of tumours, ocular angiogenic diseases, Osler-Webber syndrome, myocardial angiodysplasia, plaque neovascularisation, telangiectasia, haemophilic joints, angiodysplasia and wound stimulation of endothelial cells e.g. intestinal adhesions, atherosclerosis, scleroderma. The protein may also be useful as a birth control agent by reducing or preventing uterine vascularisation. The gene for endostatin may be isolated from cells or tissue that express high levels of endostatin, eg. tumour cells, by generating cDNA from mRNA using reverse transcriptase and then amplifying the DNA sequence.

XX SQ Sequence 182 AA:

QY Query Match 86.8%; Score 840; DB 21; Length 182;  
Best Local Similarity 85.6%; Pred. No. 7.7e-93;  
Matches 155; Conservative 15; Mismatches 11; Indels 0; Gaps 0;

DB 1 HTHQDFQPVHLVLTMPISGMRGIRGADFOCFQOARAVGLSGFRFLSRLODLYSI 60  
1 hshrdtqpvhlhvalnspisgmrgirgadfcfgqaravglagtlfrrflssrlqdllysi 60

QY 61 VRRADRGSVPIYNLKDEVSLFSGSGQLQPGARIFSGRDVLRHPAMPQKSW 120  
61 vrradtraavpivnlkdelifswsaelfsgeegplkpgariffsdgkdvlrhptwpqksw 120

DB 121 HGSDPSGRRLMESYCEWTRETTGATGQASSLLSGRLLEOKASCHNSYIVLCIENSFMT 180  
121 hgsdpsgrlrlesyctwrtettgatgqassllsgrlleokaschnsyivlcienfnt 180

QY 181 S 181  
181 a 181

DB 181 a 181

Search completed: August 28, 2001, 12:51:12  
Job time: 127 sec



---

